



Attribution–NonCommercial–NoDerivs 2.0 KOREA

You are free to :

- **Share** — copy and redistribute the material in any medium or format

Under the following terms :



Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.



NonCommercial — You may not use the material for [commercial purposes](#).



NoDerivs — If you [remix, transform, or build upon](#) the material, you may not distribute the modified material.

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

This is a human-readable summary of (and not a substitute for) the [license](#).

[Disclaimer](#) 

**A DISSERTATION FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY**

**Establishment of authentication systems  
for seven *Panax* species and a hairy root  
transformation system for *P. ginseng***

**By**

**NGUYEN VAN BINH**

**FEBRUARY, 2018**

**MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY  
DEPARTMENT OF PLANT SCIENCE  
COLLEGE OF AGRICULTURE AND LIFE SCIENCES  
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY**

**Establishment of authentication systems  
for seven *Panax* species and a hairy root  
transformation system for *P. ginseng***

**UNDER THE DIRECTION OF DR. TAE-JIN YANG  
SUBMITTED TO THE FACULTY OF THE GRADUATE  
SCHOOL OF SEOUL NATIONAL UNIVERSITY**

**BY  
NGUYEN VAN BINH**

**MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY  
DEPARTMENT OF PLANT SCIENCE**

**FEBRUARY, 2018**

**APPROVED AS QUALIFIED DISSERTATION OF NGUYEN VAN  
BINH FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**

**BY THE COMMITTEE MEMBERS**

**CHAIRMAN** \_\_\_\_\_(Seal)  
Hee-Jong Koh, Ph.D.

**VICE-CHAIRMAN** \_\_\_\_\_(Seal)  
Tae-Jin Yang, Ph.D.

**MEMBER** \_\_\_\_\_(Seal)  
Byoung-Cheorl Kang, Ph.D.

**MEMBER** \_\_\_\_\_(Seal)  
Nam-Chon Paek, Ph.D.

**MEMBER** \_\_\_\_\_(Seal)  
Beom-Seok Park, Ph.D.

# **Establishment of authentication systems for seven *Panax* species and a hairy root transformation system for *P. ginseng***

NGUYEN VAN BINH

**Department of Plant Science**

**The Graduate School of Seoul National University**

## **GENERAL ABSTRACT**

*Panax* species belong to the Araliaceae family and regarded as the king of herbal medicinal plants owing to the abundance of pharmacologically valuable ginsenosides. Among 14 reported *Panax* species, five major species including *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, and *P. vietnamensis* have been well known, and are broadly utilized as herbaceous medicinal plants in Korea, the USA, Canada, China, Japan, and Vietnam. However, their genetic diversity, evolutionary relationship, and origin remain largely unresolved. Each *Panax* species is pharmacologically and economically important, albeit with differences in efficacy and price. These lead to the fraudulent, intentional substitution or addition of a substance in a product for financial advantage. *Panax* species have been harvested mostly from the wild except four cultivated species including *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, and *P. vietnamensis*. *P. ginseng* is one of the most known *Panax* species because it contains a relatively more diverse

types of ginsenosides. However, cultivation of *P. ginseng* takes a long time (about 4 - 6 years) with extensive efforts to control quality from biotic and abiotic stresses. All these problems and limitations have affected the development of ginseng industry. To overcome these limitations and understand their genetic diversity and evolutionary relationship, I conducted a comparative analysis of whole chloroplast genomes from seven *Panax* species and established the authentication systems based on chloroplast genome sequences. Furthermore, I developed a hairy root transformation system for biomass and ginsenoside production in *P. ginseng*.

In the first chapter, I conducted a comparative analysis of whole chloroplast genome sequences of five *Panax* species including *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, and *P. vietnamensis*. I identified the number of short and large repetitive sequences, and screened large numbers of InDels and SNPs among these five species. Based on the large InDel regions, I developed fourteen practical InDel markers for authentication among these five *Panax* species, and eight of those markers were species-specific markers that successfully discriminated one unique species from the others. These markers are reliable, easily detectable, and valuable for applications in the ginseng industry as well as in related research.

In the second chapter, I completed the chloroplast genome sequences of two more basal *Panax* species (*P. stipuleanatus* and *P. trifolius*). Comparative analysis of the chloroplast genome sequences of the seven *Panax* species revealed the numbers of SNPs in protein-coding genes and whole chloroplast genome level. Phylogenetic analysis based on whole chloroplast genomes clearly showed evolutionary relationship between the seven *Panax* and their relative species in Araliaceae family. By comparing chloroplast and mitochondrial genomes, I discovered a large number of fragments from the

chloroplast genome, which is approximately 38.6%, transferred into the mitochondrial genome. I developed 18 species-specific SNP markers from the chloroplast coding sequences after eliminating intraspecies polymorphic sites, and chloroplast gene transfer regions. All these markers successfully distinguished one species from another, and can be used to authenticate all the seven *Panax* species from each other, thereby furthering efforts to protect the ginseng industry from economically motivated adulteration.

In the third or final chapter, I developed an efficient system for ginsenoside production through hairy root biomass mediated by *Agrobacterium rhizogenes* transformation. Among five transformed lines, I selected two lines which showed highest efficiency of transformation including Yunpoong cultivar and Ganghwa local landrace. PCR and RT-PCR analysis for *rol* genes in these two transformed root lines indicated that these are transgenic hairy roots induced by *A. rhizogenes* transformation. Transgenic hairy roots can grow faster than adventitious roots, and transgenic hairy root growth were not affected by auxin (IBA) application. Transgenic hairy roots produced the identical ginsenoside in comparison with those of adventitious roots. Even though total ginsenoside contents tended to be lower than those of adventitious roots, some major ginsenosides were biosynthesized in similar amounts or higher than those in adventitious roots. Bioreactor was indicated as an ideal system for large scale production of transgenic hairy roots.

This study provided valuable genetic information which can be used for future studies. The evolutionary relationship for seven *Panax* and their relative species were elucidated by whole chloroplast genomes. Two kinds of molecular markers developed in this studies can be used to authenticate all the seven *Panax* species from the others, thereby furthering efforts to protect

the ginseng industry from economically motivated adulteration. Transformation system established in this study can be applied for future related researches, and transgenic hairy roots produced here will be a valuable material resources for biomass and ginsenoside productions as well as other studies.

**Keywords:** *Panax* species, chloroplast genome, molecular markers, ginseng authentication, plant transformation, hairy root, *A. rhizogenes*, ginsenosides

**Student Number:** 2014-30833

# CONTENTS

GENERAL ABSTRACT .....	v
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vi
LIST OF ABBREVIATIONS .....	
GENERAL INTRODUCTION .....	
REFERENCES .....	
CHAPTER I .....	
Chloroplast genome derived InDel marker system to authenticate five major <i>Panax</i> species	
ABSTRACT .....	
INTRODUCTION .....	1
METERIAL AND METHODS .....	1
RESULTS .....	1
Structure of complete chloroplast genomes of five <i>Panax</i> species .....	1
Divergence of chloroplast genomes among <i>Panax</i> species .....	2
Development of InDel markers to identify <i>Panax</i> species .....	2
DISCUSSION .....	2
REFERENCES .....	3
CHAPTER II .....	3
Comprehensive chloroplast genomes analysis of seven <i>Panax</i> species and development of an authentication system based on species-unique SNP markers derived from chloroplast genomes	
ABSTRACT .....	3
INTRODUCTION .....	4
METERIAL AND METHODS .....	4



<b>RESULTS</b>	4
Characteristics of the complete chloroplast genomes from seven <i>Panax</i> species	4
Phylogenomic analysis for 19 complete chloroplast genomes of Araliaceae species	4
SNPs in chloroplast genomes of seven <i>Panax</i> species	5
Characterization of chloroplast genome transfer into the mitochondrial genome	5
Identification of species-specific SNP markers for authentication of the seven <i>Panax</i> species	5
<b>DISCUSSION</b>	6
<b>REFERENCES</b>	6
<b>CHAPTER III</b>	7
<b>Optimal condition and genotypes for transgenic hairy root induction and metabolite production in <i>P. ginseng</i></b>	
<b>ABSTRACT</b>	7
<b>INTRODUCTION</b>	7
<b>MATERIAL AND METHODS</b>	7
<b>RESULTS</b>	8
Efficiency of <i>A. rhizogenes</i> -mediated transformation among five <i>P. ginseng</i> accessions	8
Molecular analysis of transgenic hairy roots	8
Comparison of hairy root growth rate in liquid shaking culture	8
Ginsenoside profiling in hairy roots compared to adventitious roots	9
Biomass production of hairy roots in bioreactor	9
<b>DISCUSSION</b>	9
<b>REFERENCES</b>	10
<b>ABSTRACT IN KOREAN</b>	10

## LIST OF TABLES

<b>Table 1-1.</b> SSRs in the five <i>Panax</i> chloroplast genomes.....	17
<b>Table 1-2.</b> Numbers and ratios of nucleotide substitutions and InDels among chloroplast genomes of five <i>Panax</i> species .....	22
<b>Table 1-3.</b> Molecular markers developed to authenticate <i>Panax</i> species.....	25
<b>Table 1-4.</b> Marker combinations for each <i>Panax</i> species.....	26
<b>Table 2-1.</b> Chloroplast genome sequences used in this study.....	47
<b>Table 2-2.</b> Numbers of SNPs among seven <i>Panax</i> chloroplast genomes.....	52
<b>Table 2-3.</b> Details for dCAPS markers developed to authenticate <i>Panax</i> species .....	57
<b>Table 3-1.</b> Primer information for PCR and RT-PCR analysis.....	85

## LIST OF FIGURES

<b>Figure 1-1.</b> Complete chloroplast genomes of five <i>Panax</i> species .....	18
<b>Figure 1-2.</b> Repeat structure analysis of the five <i>Panax</i> species chloroplast genomes .....	19
<b>Figure 1-3.</b> Comparison of chloroplast genome sequences of five <i>Panax</i> species .....	23
<b>Figure 1-4.</b> Validation of 14 molecular markers derived from InDel regions of five <i>Panax</i> chloroplast genomes.....	27
<b>Figure 2-1.</b> Complete chloroplast genomes of seven <i>Panax</i> species .....	48
<b>Figure 2-2.</b> ML phylogenetic tree of <i>P. stipuleanatus</i> and <i>P. trifolius</i> with 17 related species in the Araliaceae family based on entire chloroplast genome sequences .....	50
<b>Figure 2-3.</b> Single nucleotide polymorphic sites in 79 CDS genes of 7 <i>Panax</i> species .....	53
<b>Figure 2-4.</b> Schematic representation of gene transfer between chloroplast and mitochondrial genome of <i>Panax</i> species .....	55
<b>Figure 2-5.</b> Validation of 18 dCAPS markers derived from CDS SNP regions of seven <i>Panax</i> chloroplast genomes.....	59
<b>Figure 3-1.</b> Efficiency of <i>Agrobacterium rhizogenes</i> transformation of five <i>P. ginseng</i> accessions .....	83
<b>Figure 3-2.</b> Analysis in transgenic hairy roots of ginseng.....	86
<b>Figure 3-3.</b> GUS histochemical analysis of transgenic hairy roots of ginseng transformed by <i>A. rhizogenes</i> R100 (pBI121).....	87
<b>Figure 3-4.</b> Growth rate of transgenic hairy roots in liquid medium .....	89
<b>Figure 3-5.</b> Morphology of adventitious roots and transgenic hairy roots in liquid medium.....	90

<b>Figure 3-6.</b> UPLC-QTOP/MS chromatogram of adventitious roots and transgenic hairy roots.....	92
<b>Figure 3-7.</b> Quantitative analysis of 13 major ginsenosides in adventitious roots and transgenic hairy roots.....	93
<b>Figure 3-8.</b> Biomass and ginsenoside production of transgenic hairy roots in bioreactor.....	95

## LIST OF ABBREVIATIONS

WGS	Whole genome sequence
NGS	Next generation sequencing
dnaLCW	<i>de novo</i> assembly using low coverage of WGS
nrDNA	Nuclear ribosomal DNA
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
LSC	Large single copy
SSC	Small single copy
IR	Inverted repeat
PCR	Polymerase chain reaction
InDel	Insertion or Deletion
dCAPS	Derived cleaved amplified polymorphic sequences
CDS	Coding sequence
IGS	Intergenic space
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
TRs	Tandem repeats
EMA	Economically motivated adulteration
WGD	Whole genome duplication
SH	Schenk and Hildebrandt
MS	Murashige and Skoog 1962
IBA	Indole-3-butyric acid

## GENERAL INTRODUCTION

The genus *Panax* belongs to the Araliaceae family and are widely distributed from high-altitude freeze-free area of Eastern Himalayas, Hoang Lien Son and Annamite mountain ranges to freezing winter area of Northeastern Asia and North America. To date, more than 200 ginsenosides, a triterpene saponin specifically biosynthesized in *Panax* species, and non-saponin constituents have been isolated from ginseng (Zhang *et al.* 2006). Currently, ginseng is becoming one of the most important national agricultural commodities not only in Asian countries such as Korea, China, and Vietnam, but also in Russia, Canada and the USA. Of the 14 known species in the *Panax* genus, five species, *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, and *P. vietnamensis*, have been known well and utilized as expensive herbal medicines in Korea, the USA, Canada, China, Japan, and Vietnam. The remaining species still have limited genetic and pharmaceutical information. Each *Panax* species have their own pharmacological and economical value. Owing to the high value of ginseng, it has become a target of economically motivated adulterations (EMAs). These EMA products have resulted in serious unintended public health consequences and illustrated gaps in the regulatory and quality assurance system. Moreover, the global demand for herbal medicines is steadily increasing due to reinsured interest of consumers in natural products (Chen *et al.* 2016; Hishe *et al.* 2016; Verma and Singh 2008), and ginseng is also one of the best concern because of high pharmacological value. However, most *Panax* species are harvested from the wild except for four cultivated species including *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, and *P. vietnamensis*. In addition, cultivation of ginseng is a big challenge because of their need of special growing conditions. Altogether, these contribute to the

limitations in the development of the ginseng industry, thus the method to authenticate ginseng products and a new material resource for ginseng biomass are in high demand.

Many methods have been applied to authenticate *Panax* species. These include the more traditional morphology-based, chemical analysis-based, and DNA barcoding methods. However, traditional methods are not precise enough because of the similar morphological appearances between species or the different morphological character of intraspecies related to growing conditions. Moreover, many commercial ginseng products are sold in a processed form, such as red ginseng, ginseng powder, liquid extracts, pellets, shredded slices, or even tea, which cannot be authenticated by traditional methods. Authentication based on ginsenoside profiling methods are normally expensive and difficult to utilize in high-throughput. In addition, these applications are limited because ginsenosides are secondary metabolites and the accumulation of ginsenosides vary in quantity between different tissues (such as roots, leaves, stems, flower buds and berries) (Oh *et al.* 2014; Shi *et al.* 2007), cultivars (Lee *et al.* 2017), age (Shi *et al.* 2007; Xiao *et al.* 2015), and environmental conditions (Jiang *et al.* 2016; Kim *et al.* 2014). Additionally, storage conditions and manufacturing processes also affect the secondary metabolite accumulation in ginseng. Sequence-based DNA markers have shown the advantages and power in species identification with high accuracy, simplicity, and time- and cost-efficiency. Various DNA markers have been applied for authentication of *Panax* species including nuclear and chloroplast genome-derived markers (Artyukova *et al.* 2000; Choi *et al.* 2011; Jung *et al.* 2014; Kim *et al.* 2013; Kim *et al.* 2015; Kim *et al.* 2012; Ma *et al.* 2007). However, thus far, the application of these markers is still limited and almost used to identify only at the intraspecies level of *P. ginseng*, and just few *Panax* species at the interspecies level.

Various strategies and approaches have been developed using cell and organ plant cultures via biomass improvement for production of secondary metabolites in plants (Murthy *et al.* 2014). Similar techniques have also been developed to produce ginseng biomass in *in vitro* system for ginsenoside production, such as producing biomass of callus, suspension, and adventitious roots (Ali *et al.* 2006; Furuya *et al.* 1984; Langhansova *et al.* 2005; Nguyen and Paek 2010; Woo *et al.* 2004; Yu *et al.* 2002). In addition, many propagation systems have been invented for biomass production, such as bioreactor, liquid shaking culture, and temporary immersion culture (Baque *et al.* 2012; Paek *et al.* 2005; Vaněk *et al.* 2005). However, these methods are still practically limited due to low growth rate of plant tissues, and requirement of supplemental plant hormones for their growth. This led to increased cost of ginsenoside production, thus a new efficient material resource that can solve these limitations will help to develop ginseng industry.

Overall, in this study, I tried to comprehensively understand the genetic diversity and evolutionary relationship between *Panax* species, and based on that, I developed authentication systems for *Panax* species, which can be used to discriminate among these seven *Panax* species and others. These markers are valuable tools for genetic studies in ginseng and can be applied towards protecting the ginseng industry from economically motivated adulteration. Moreover, I also developed an efficient method for the development of transgenic hairy roots which will provide a new material resource for large scale biomass and ginsenoside productions in *P. ginseng*, and a biotechnological tool for related studies in ginseng.



## REFERENCES

- Ali MB, Yu KW, Hahn EJ, Paek KY (2006) Methyl jasmonate and salicylic acid elicitation induces ginsenosides accumulation, enzymatic and non-enzymatic antioxidant in suspension culture *Panax ginseng* roots in bioreactors. *Plant Cell Reports* 25:613-620
- Artyukova E, Kozyrenko M, Reunova G, Muzarok T, Zhuravlev YN (2000) RAPD analysis of genome variability of planted ginseng, *Panax ginseng*. *Molecular Biology* 34:297-302
- Baque MA, Moh SH, Lee EJ, Zhong JJ, Paek KY (2012) Production of biomass and useful compounds from adventitious roots of high-value added medicinal plants using bioreactor. *Biotechnology Advances* 30:1255-1267
- Chen SL, Yu H, Luo HM, Wu Q, Li CF, Steinmetz A (2016) Conservation and sustainable use of medicinal plants: problems, progress, and prospects. *Chinese Medicine* 11:37
- Choi HI, Kim NH, Kim JH, Choi BS, Ahn IO, Lee JS, Yang TJ (2011) Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. *Journal of Ginseng Research* 35:399
- Furuya T, Yoshikawa T, Orihara Y, Oda H (1984) Studies of the culture conditions for *Panax ginseng* cells in jar fermentors. *Journal of Natural Products* 47:70-75
- Hishe M, Asfaw Z, Giday M (2016) Review on value chain analysis of medicinal plants and the associated challenges. *Journal of Medicinal Plants Studies* 4:45-55
- Jiang M, Liu J, Quan X, Quan L, Wu S (2016) Different chilling stresses stimulated the accumulation of different types of ginsenosides in *Panax ginseng* cells. *Acta Physiol Plant* 38:210

- Jung J, Kim KH, Yang K, Bang KH, Yang TJ (2014) Practical application of DNA markers for high-throughput authentication of *Panax ginseng* and *Panax quinquefolius* from commercial ginseng products. *Journal of Ginseng Research* 38:123-129
- Kim JH, Jung JY, Choi HI, Kim NH, Park JY, Lee Y, Yang TJ (2013) Diversity and evolution of major *Panax* species revealed by scanning the entire chloroplast intergenic spacer sequences. *Genetic Resources and Crop Evolution* 60:413-425
- Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, Park HS, Yang TJ (2015) Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS One* 10:e0117159.
- Kim NH, Choi HI, Ahn IO, Yang TJ (2012) EST-SSR marker sets for practical authentication of all nine registered ginseng cultivars in Korea. *Journal of Ginseng Research* 36:298
- Kim YJ, Jeon JN, Jang MG, Oh JY, Kwon WS, Jung SK, Yang DC (2014) Ginsenoside profiles and related gene expression during foliation in *Panax ginseng* Meyer. *Journal of Ginseng Research* 38:66-72
- Langhansova L, Marsik P, Vanek T (2005) Production of saponins from *Panax ginseng* suspension and adventitious root cultures. *Biologia Plantarum* 49:463-465
- Lee YS, Park HS, Lee DK, Jayakodi M, Kim NH, Koo HJ, Lee SC, Kim YJ, Kwon SW, Yang TJ (2017) Integrated transcriptomic and metabolomic analysis of five *Panax ginseng* cultivars reveals the dynamics of ginsenoside biosynthesis. *Frontiers in Plant Science* 8:1048
- Ma KH, Dixit A, Kim YC, Lee DY, Kim TS, Cho EG, Park YJ (2007) Development and characterization of new microsatellite markers for ginseng (*Panax ginseng* CA Meyer). *Conservation Genetics* 8:1507-1509

- Murthy HN, Lee EJ, Paek KY (2014) Production of secondary metabolites from cell and organ cultures: strategies and approaches for biomass improvement and metabolite accumulation. *Plant Cell, Tissue and Organ Culture (PCTOC)* 118:1-16
- Nguyen TT, Paek KY (2010) Cell suspension culture *Panax ginseng* CA Meyer: Role of plant growth regulators and medium composition on biomass and ginsenoside production. *Natural Sciences and Technology* 26:191-196
- Oh JY, Kim YJ, Jang MG, Joo SC, Kwon WS, Kim SY, Jung SK, Yang DC (2014) Investigation of ginsenosides in different tissues after elicitor treatment in *Panax ginseng*. *Journal of Ginseng Research* 38:270-277
- Paek K, Chakrabarty D, Hahn E (2005) Application of bioreactor systems for large scale production of horticultural and medicinal plants. *Plant Cell, Tissue Organ Culture* 81:287-300
- Shi W, Wang Y, Li J, Zhang H, Ding L (2007) Investigation of ginsenosides in different parts and ages of *Panax ginseng*. *Food Chemistry* 102:664-668
- Vaněk T, Langhansová L, Maršík P (2005) Cultivation of root cultures of *Panax ginseng* in different bioreactors and in temporary immersion-Comparison of growth and saponin production. In: Hvoslef-Eide A, Preil W (Eds) *Liquid culture systems for in vitro plant propagation*, Springer, The Netherlands, pp 539–547
- Verma S, Singh S (2008) Current and future status of herbal medicines. *Veterinary World* 1:347-350
- Woo SS, Song JS, Lee JY, In DS, Chung HJ, Liu JR, Choi DW (2004) Selection of high ginsenoside producing ginseng hairy root lines using targeted metabolic analysis. *Phytochemistry* 65:2751-2761

- Xiao D, Yue H, Xiu Y, Sun X, Wang Y, Liu S (2015) Accumulation characteristics and correlation analysis of five ginsenosides with different cultivation ages from different regions. *Journal of Ginseng Research* 39:338-344
- Yu KW, Gao W, Hahn EJ, Paek KY (2002) Jasmonic acid improves ginsenoside accumulation in adventitious root culture of *Panax ginseng* CA Meyer. *Biochemical Engineering Journal* 11:211-215
- Zhang J, Chui D, Chen C, Liu G (2006) The chemistry, metabolism and biological activities of ginseng. Beijing: Chemical Industry Press

# **CHAPTER I**

**Chloroplast genome derived InDel marker system to  
authenticate five major *Panax* species**

## ABSTRACT

Ginseng represents a set of high-value medicinal plants of different species: *Panax ginseng* (Asian ginseng), *P. quinquefolius* (American ginseng), *P. notoginseng* (Chinese ginseng), *P. japonicus* (Bamboo ginseng), and *P. vietnamensis* (Vietnamese ginseng). Each species is pharmacologically and economically important, with differences in efficacy and price. Accordingly, an authentication system is needed to combat economically motivated adulteration of *Panax* products. I conducted comparative analysis of the chloroplast genome sequences of these five species, identifying 34–124 InDels and 141–560 SNPs. Fourteen InDel markers were developed to authenticate the *Panax* species. Among these, eight were species-unique markers that successfully differentiated one species from the others. I generated at least one species-unique marker for each of the five species, and any of the species can be authenticated by selection among these markers. The markers are reliable, easily detectable, and valuable for applications in the ginseng industry as well as in related research.

**Keywords:** *Panax* species, Chloroplast genome, Molecular markers, Ginseng authentication

## INTRODUCTION

The *Panax* genus belongs to the Araliaceae family and contains many important medicinal species, collectively called ‘ginseng’. Of the 14 known species in the *Panax* genus, five species, *Panax ginseng* (Asian ginseng), *P. quinquefolius* (American ginseng), *P. notoginseng* (Sanchi ginseng; Chinese ginseng), *P. japonicus* (Bamboo ginseng), and *P. vietnamensis* (Vietnamese ginseng), are broadly utilized in Korea, the USA, China, Japan, and Vietnam. Each species is well-known as a traditional medicinal plant in oriental countries, and species such as *P. ginseng*, *P. quinquefolius*, and *P. notoginseng* contain protopanaxadiol-type and protopanaxatriol-type saponins (Zhu *et al.* 2004), while other species like *P. japonicus* and *P. vietnamensis* contain high quantities of oleanolic acid-type and ocotillol-type saponins, respectively (Yamasaki 2000; Zhu *et al.* 2004).

The high pharmacological and economical value of ginseng means that many economically motivated adulterations (EMAs) of ginseng products have been developed by substituting morphologically similar plant roots, or by mixing different species. Traditionally, the authentication of herb plants was based on morphological and histological inspection. However, these traditional methods are unable to authenticate some *Panax* species because of their very similar morphological appearances, especially in terms of root shape. For example, *P. ginseng* and *P. quinquefolius*, and *P. japonicus* and *P. vietnamensis* cannot easily be distinguished from each other. Moreover, many commercial ginseng products are sold in a processed form, such as red ginseng, ginseng powder, liquid extracts, pellets, shredded slices, or even tea, which cannot be authenticated by traditional methods. Ginsenoside profiling methods have been developed to authenticate ginseng samples (Chan *et al.* 2000). However, the effects of factors such as growth conditions, developmental stage, internal metabolism, storage conditions, and

manufacturing processes on secondary metabolite accumulation in ginseng limits the application of such chemical analyses (Ngan *et al.* 1999). Chemical methods are also expensive and difficult to utilize in high-throughput analysis. Therefore, reliable and practical methods to authenticate ginseng are in high demand.

The chloroplast is a plant-specific organelle containing the entire machinery required for photosynthesis and carbon fixation. Chloroplast genomes are generally highly conserved across land plants at the gene level, with a quadripartite structure comprising two inverted repeat blocks (IRs), one large single copy (LSC) region, and one small single copy (SSC) region. As a result of interspecies sequence divergence and intraspecies sequence conservation, the chloroplast genome is valuable for taxonomic classification and phylogeny reconstruction (Huang *et al.* 2014; Kim *et al.* 2015a). Lack of recombination, low nucleotide substitution rates, and usually uniparental inheritance make chloroplast genomes valuable sources of genetic markers for phylogenetic analysis and species identification (Li *et al.* 2013; Wolfe *et al.* 1987). Chloroplast sequences such as *atpF-atpH*, *matK*, *psbK-psbI*, *rbcL*, *ropC1*, *rpoB*, and *trnH-psbA* are commonly used as DNA barcodes for plants (Dong *et al.* 2012; Elansary *et al.* 2017; Hollingsworth *et al.* 2009). In some cases, these sequences were highly efficient for species identification and phylogenetic studies, but they showed low variation in closely related species (Dong *et al.* 2012).

Recently, DNA markers have been developed to authenticate ginseng, including random amplified polymorphic DNA (RAPD) (Artyukova *et al.* 2000), microsatellite (Ma *et al.* 2007), and expressed sequence tag-simple sequence repeat (EST-SSR) (Choi *et al.* 2011; Kim *et al.* 2012) markers, as well as single nucleotide polymorphism (SNP) and insertion and deletion (InDel) markers derived from chloroplast sequences (Jung *et al.* 2014; Kim



*et al.* 2013; Kim *et al.* 2015a). However, thus far, these markers have been used to identify only *P. ginseng* cultivars at the intraspecies level, and just a few *Panax* species at the interspecies level.

Recently, we developed an efficient method to obtain complete chloroplast genome and nuclear ribosomal DNA (nrDNA) by *de novo* assembly using low-coverage whole-genome shotgun next-generation sequencing (dnaLCW) (Kim *et al.* 2015b). Using this method, we obtained complete chloroplast genomes and nrDNA for the five *Panax* species: *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, and *P. vietnamensis* (Kim *et al.* 2015a; Kim *et al.* 2017). In the present study, I comparatively analyzed these five chloroplast genome sequences and developed credible chloroplast genome-derived InDel markers to authenticate these *Panax* species. These markers are valuable tools for the further study of genetic diversity in the *Panax* genus. They also may be used to support the ginseng industry, which depends on a number of *Panax* species, for example, *P. ginseng* in Korea, China, and Japan, *P. quinquefolius* in the USA, Canada, and China, *P. notoginseng* in China, and *P. vietnamensis* in Vietnam.

## MATERIAL AND METHODS

### Plant materials and Genomic DNA extraction

*P. ginseng* (cultivars ‘Chunpoong’ (CP) and ‘Yunpoong’ (YP)) and *P. quinquefolius* plants were collected from the ginseng farm at Seoul National University in Suwon, Korea. *P. notoginseng* and *P. japonicus* plants were collected from Dafang County, Guizhou Province, and Enshi County, Hubei Province, China, respectively. *P. vietnamensis* plants were collected from Ngoc Linh Mountain, Kon Tum Province, Vietnam. Individual leaves and roots of plants from each species were harvested and stored at  $-70^{\circ}\text{C}$  until use. Total genomic DNA was isolated using a modified cetyltrimethylammonium bromide (CTAB) method (Allen *et al.* 2006). The quality and quantity of extracted genomic DNA was measured using a UV-spectrophotometer.

### Comparative analysis and characterization of SSRs and large sequence repeats

The chloroplast genome sequences of *P. ginseng* cv. CP (KM088019), *P. ginseng* cv. YP (KM088020), *P. quinquefolius* (KM088018), *P. notoginseng* (KP036468), *P. japonicus* (KP036469), and *P. vietnamensis* (KP036471) were obtained from our previous studies (Kim *et al.* 2015a; Kim *et al.* 2017). MAFFT (<http://mafft.cbrc.jp/alignment/server/>), MEGA 6 (Tamura *et al.* 2013), and mVISTA (<http://genome.lbl.gov/vista/mvista/submit.shtml>) programs were used to compare these sequences.

SSRs were predicted using MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) (Thiel *et al.* 2003) with the parameters set to  $\geq 10$  repeat units for mononucleotide SSRs,  $\geq 5$  repeat units for dinucleotide SSRs,  $\geq 4$  repeat units for trinucleotide SSRs, and  $\geq 3$  repeat units for tetranucleotide, pentanucleotide, and hexanucleotide SSRs.

The program REPuter (Kurtz *et al.* 2001) was used to identify the number and location of repeat sequences, including tandem, dispersed, reverse, and palindromic repeats within chloroplast genomes. For all repeat types, the following constraints were set to a minimum repeat size of 15 bp, and 90% minimum cut-off identity between two copies. Overlapping repeats were merged into one repeat motif whenever possible.

### **Development and validation of InDel markers**

To validate intraspecies and interspecies polymorphism in the chloroplast genomes, and develop DNA markers to authenticate the five major *Panax* species, specific primers were designed based on InDel polymorphic sites found in *Panax* chloroplast genomes, including *P. ginseng* cv. YP (KM088020) as an intraspecies level. Primer pairs were designed using the Primer3 program (<http://bioinfo.ut.ee/primer3-0.4.0/>).

The polymerase chain reaction (PCR) was performed in a 25 µl reaction mixture containing 20 ng of DNA template, 5 pmol of each primer, 1.25 mM deoxynucleotide triphosphate (dNTP), 1.25 units of *Taq* DNA polymerase (Inclone, Korea), and 2.5 µl of 10× reaction buffer. The PCR reaction was performed in thermocyclers using the following cycling parameters: 94°C (5 min); 35 cycles of 94°C (30 s), 54–58°C (30 s); 72°C (30 s), then 72°C (7 min). PCR products were visualized on agarose gels (1.5–3.0%) after staining with ethidium bromide, and/or analyzed by capillary electrophoresis using a fragment analyzer (Advanced Analytical Technologies Inc., USA).

## RESULTS

### Structure of complete chloroplast genomes of five *Panax* species

Complete chloroplast genomes of five *Panax* species were obtained by dnaLCW and reported in our previous studies (Kim *et al.* 2015a; Kim *et al.* 2017). Lengths of the chloroplast genomes ranged from 155,993 bp (*P. vietnamensis*) to 156,466 bp (*P. notoginseng*). The order, content, and orientation of genes were highly conserved among these five chloroplast genomes, comprising 79 protein-coding genes, 30 tRNA genes, and 4 rRNA genes in common. Each chloroplast genome had the same quadripartite structure with an LSC, an SSC, and two IR regions (Fig. 1-1).

Variations in the copy numbers of SSRs were identified among the five *Panax* chloroplast genomes. The longest SSRs were 18 nucleotides in length, and the most abundant nucleotides in SSRs were A and T (Table 1-1). *P. vietnamensis* and *P. notoginseng* both contained 42 SSRs; lower than *P. japonicus* (45), and higher than *P. ginseng* cv. CP (38) and *P. quinquefolius* (40) (Table 1-1). The *P. japonicus* chloroplast genome has the highest number of homopolymers (24), but no hexapolymers. *P. vietnamensis*, *P. notoginseng*, and *P. ginseng* cv. CP all had 6 dipolymers; lower than *P. japonicus* and *P. quinquefolius* (7). *P. notoginseng* had 4 tripolymers; more than each of the other *Panax* species (3). *P. vietnamensis* had 10 tetrapolymers; more than the other *Panax* species (8). *P. japonicus* and *P. notoginseng* had 3 pentapolymers each, while the other *Panax* species contained 2 (Table 1-1).

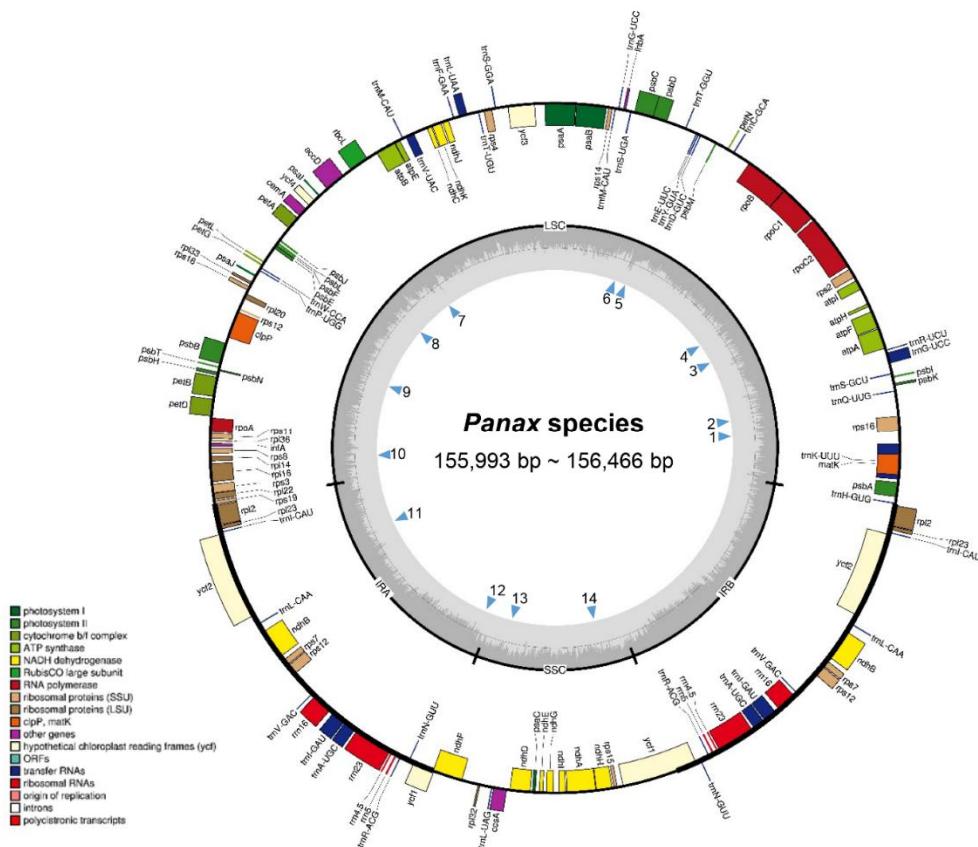
For comparative analysis, repetitive sequences were grouped into four categories: tandem, dispersed, palindromic, and reverse. To avoid redundancy, repeat sequence analysis of each chloroplast genome was carried out with a single IR region. A total of 45–50 repetitive sequences

were identified in each of the five *Panax* chloroplast genomes (Fig. 1-2B), including dispersed repeats (44.26%), palindromic repeats (29.79%), tandem repeats (22.13%), and reverse repeats (3.83%). Most repeats were located in intergenic spaces (IGS) (53.83%) and coding sequence (CDS) regions (38.94%); a small number of repeats (7.23%) were found in intron regions (Fig. 1-2C). Across the five chloroplast genomes, repeat lengths ranged from 17 bp to 67 bp (Fig. 1-2A). The 67-bp tandem repeat found in *P. japonicus* was the longest repeat (Fig. 1-2A). Palindromic and reverse repeats ranged from 17–35 bp and 17–28 bp, respectively (Fig. 1-2A). Comparing the length and location, 32 repeats were shared by all five *Panax* species, and 5 repeats were present in four chloroplast genomes. *P. notoginseng* had the most unique repeats (9), followed by *P. ginseng* and *P. vietnamensis* (6), *P. japonicus* (3), and *P. quinquefolius* (1) (Fig. 1-2D).

**Table 1-1.** SSRs in the five *Panax* chloroplast genomes.

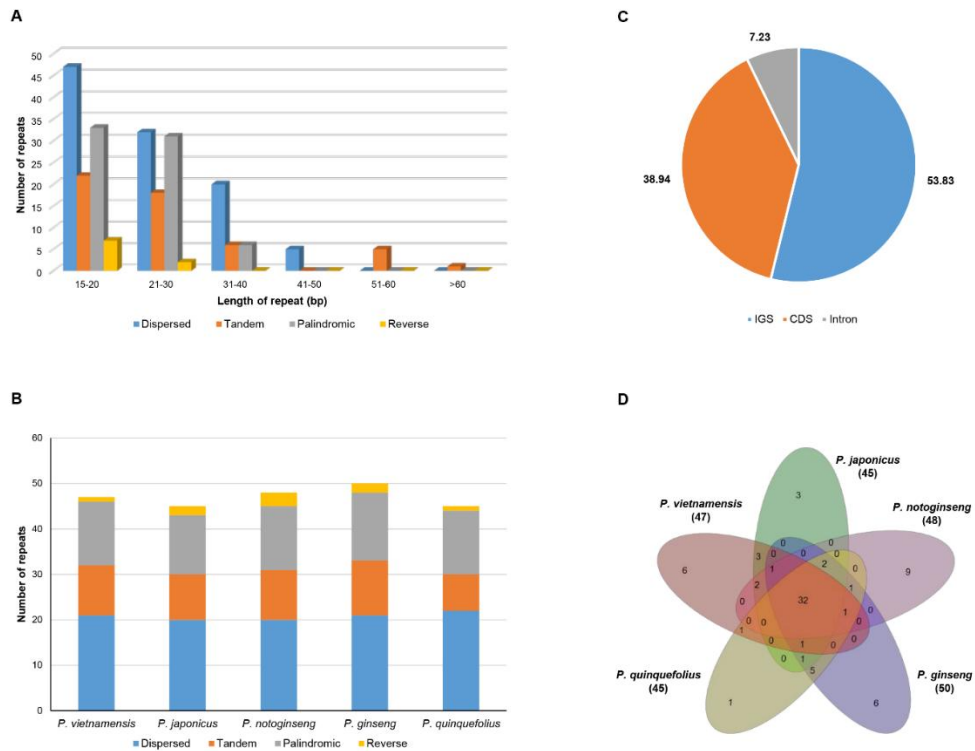
Motif	Repeat units	Number of SSRs				
		PgCP	Pq	Pn	Pj	Pv
Mononucleotide	A/T	14	16	19	17	17
	C/G	4	3	1	7	3
Dinucleotide	AT/AT	6	7	6	7	6
	AAG/CTT	1	1	1	1	1
Trinucleotide	AAT/ATT	1	1	3	2	2
	AGC/CTG	1	1	-	-	-
	AAAG/CTTT	3	3	3	3	3
Tetranucleotide	AAAT/ATTT	2	2	2	2	4
	AATT/AATT	1	1	1	1	1
	ACCT/AGGT	2	2	2	2	2
Pentanucleotide	AATCT/AGATT	2	2	2	2	2
	AAAAT/ATTTT	-	-	1	1	-
Hexanucleotide	AGATAT/ATATCT	-	-	-	-	1
	ACTATG/AGTCAT	1	1	1	-	-
Total SSRs		38	40	42	45	42

**Abbreviations:** PgCP, *P. ginseng* cv. CP; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*; SSR, simple sequence repeat.



**Figure 1-1. Complete chloroplast genomes of five *Panax* species.**

Colored boxes show conserved chloroplast genes, classified based on product function. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. Dashed area in the inner circle indicates the GC content of the chloroplast genome; blue triangles indicate the positions of 14 markers (gcpm1–gcpm14) on the chloroplast genomes.



**Figure 1-2. Repeat structure analysis of the five *Panax* species chloroplast genomes.**

(A) Histogram showing the frequency of repeats by length in the five *Panax* chloroplast genomes. (B) Histogram showing the number of four repeat types in each *Panax* chloroplast genome. (C) Location of the 235 repeats on the five *Panax* species chloroplast genomes. (D) Venn diagram showing the repeats shared among the five *Panax* species.



### **Divergence of chloroplast genomes among *Panax* species**

To investigate chloroplast genome divergence among the five *Panax* species, multiple alignments of six chloroplast genomes were performed with a single IR region. The sequence identity of six chloroplast genomes was plotted using the mVISTA program, using the *Panax ginseng* cv. CP annotation as the reference (Fig. 1-3). Two cultivating varieties of ginseng, CP and YP, were almost identical with 3 InDels and 1 SNP overall (Table 1-2). Five *Panax* species also showed high sequence similarity with each other (98.9%) (Fig. 1-3), suggesting that *Panax* chloroplast genomes are well conserved. Our results are consistent with earlier research on Araliaceae chloroplast genomes (Li *et al.* 2013). As expected, coding regions are more conserved than non-coding regions. The most highly conserved chloroplast genome coding regions are the four rRNA and 30 tRNA genes (Fig. 1-3). The most divergent coding region is the *ycf1* gene, which has low sequence identity because of various InDels and repeat sequences (Fig. 1-3). This has been reported in previous research on other chloroplast genomes (Li *et al.* 2013; Nie *et al.* 2012).

Although chloroplast genome diversity at the intraspecies level is low, abundant polymorphism was identified between species. At the interspecies level, the number of SNPs ranged between 141 (*P. ginseng* versus *P. quinquefolius*) and 560 (*P. ginseng* versus *P. vietnamensis*), and the number of InDels ranged between 34 (*P. ginseng* versus *P. quinquefolius*) and 124 (*P. notoginseng* versus *P. vietnamensis*) (Table 1-2). A/T SNP substitutions were more frequent than other types, in agreement with previous studies (Huang *et al.* 2014; Xu *et al.* 2002). Fewer substitutions and InDels were found between *P. ginseng* and *P. quinquefolius* than between the other three *Panax* species. The ratios of nucleotide substitution events to InDel events (S/I) for different pairwise comparisons between species ranged between

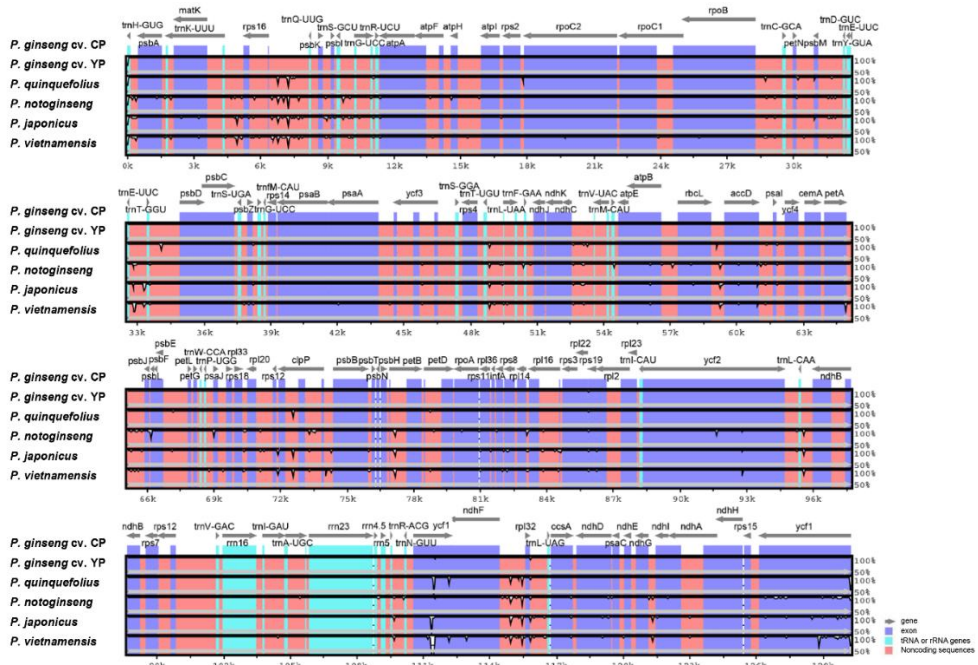
4.06 (*P. ginseng* versus *P. quinquefolius*) and 5.64 (*P. quinquefolius* versus *P. japonicus*) (Table 1-2). S/I ratios are thought to increase with divergence time between genomes (Chen *et al.* 2009). Our results indicate that *P. ginseng* is more closely related to *P. quinquefolius* (S/I = 4.06), and that *P. quinquefolius* is more highly divergent from *P. japonicus* (S/I = 5.64), compared to others.

**Table 1-2.** Numbers and ratios of nucleotide substitutions and InDels among chloroplast genomes of five *Panax* species.

	<b>PgCP</b>	<b>PgYP</b>	<b>Pq</b>	<b>Pn</b>	<b>Pj</b>	<b>Pv</b>
<b>PgCP</b>	/	1	141	480	511	559
<b>PgYP</b>	3 (0.33)	/	142	481	512	560
<b>Pq</b>	34 (4.15)	35 (4.06)	/	509	508	542
<b>Pn</b>	101 (4.75)	98 (4.91)	110 (4.63)	/	484	548
<b>Pj</b>	92 (5.55)	93 (5.51)	90 (5.64)	111 (4.36)	/	331
<b>Pv</b>	102 (5.48)	103 (5.44)	104 (5.21)	124 (4.42)	74 (4.47)	/

**Note:** The upper triangle shows the total nucleotide substitutions, while the lower triangle indicates the number of InDels. Ratios of nucleotide substitutions to InDels (S/I) are given in brackets.

Abbreviations: PgCP, *Panax ginseng* cv. CP; PgYP, *P. ginseng* cv. YP; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicas*; Pv, *P. vietnamensis*.



**Figure 1-3. Comparison of chloroplast genome sequences of five *Panax* species.**

Pair-wise comparison of chloroplast genomes between *Panax* species using the mVISTA program with *P. ginseng* cv. CP as the reference. Genome regions are color-coded as protein coding (purple), rRNA or tRNA coding genes (blue), and noncoding sequences (pink).

## Development of InDel markers to identify *Panax* species

Based on chloroplast genome sequence alignment, the 14 most InDel-variable loci were selected to develop 14 potentially discriminate markers (Table 1-3; Fig. 1-1). Each of these 14 markers were successfully amplified by PCR, and each showed the expected polymorphic band sizes. Eight of these 14 markers had unique amplicon sizes specific to different *Panax* species (Table 1-4).

The marker gcpm2 was specific to both *P. ginseng* cultivars (CP and YP), and was derived from a 33-bp tandem repeat (TR) in the *rps16-trnQ*-UUG region (Table 1-4; Fig. 1-4B). The *P. notoginseng*-specific markers gcpm3, gcpm8, and gcpm10 were derived from a 25-bp TR in the *atpH-atpI* region, a 38-bp TR in the *petA-psbJ* region, and a 25-bp TR in the *rpl14-rpl16* region, respectively (Table 1-4; Fig. 1-4C, H, K). The gcpm4 marker was derived from a 23-bp TR in the *rps2-rpoC2* region, and was specific to *P. quinquefolius* (Table 1-4; Fig. 1-4D). The marker gcpm6 was derived from a 67-bp TR and a 19-bp InDel in the *trnE-trnT* region, and was specific to *P. japonicus* (Table 1-4; Fig. 1-4F). Finally, the markers gcpm9 and gcpm14 were derived from a 25-bp TR and a 6-bp SSR in the *clpP-psbB* region, and a 30-bp InDel in *ycfI*, respectively, and were specific to *P. vietnamensis* (Table 1-4; Fig. 1-4I, O).

Validation results revealed that six markers, gcpm1, 5, 7, 11, 12, and 13, were able to identify more than two species, of which gcpm12 (derived from a 57-bp TR in the *ycfI* gene) was the most variable (Table 1-4; Fig. 1-4M). In addition, gcpm12 also distinguished between the two *P. ginseng* cultivars (Table 1-4; Fig. 1-4M).

**Table 1-3.** Molecular markers developed to authenticate *Panax* species.

Marker name	Forward (F)/reverse (R) primers	Melting temperature (°C)	Expected PCR product size (bp)					
			PgCP	PgYP	Pq	Pn	Pj	Pv
gcpm1	F: TGGGGGAATAAATAAGGAAGA A <sup>a</sup>	54.7	392	392	393	391	416	430
	R: TTAAAGAAGGCGGAGGTTTT <sup>a</sup>	54.0						
gcpm2	F: TGGAAGGCTGTTGTCACTG <sup>b</sup>	53.2	194	194	161	160	161	161
	R: TGCCAAGATTGCAGAAAGATT <sup>b</sup>	53.8						
gcpm3	F: TGGTCAATCGCTGAAGAAAA <sup>a</sup>	53.2	449	449	449	474	449	449
	R: CCGCGGCTTATATAGGTGAA <sup>a</sup>	57.3						
gcpm4	F: CTCTTCCAAATTGATGTTCCAA <sup>a</sup>	56.6	295	295	318	295	295	295
	R: TCCATGATACACCAGAACAATC A <sup>a</sup>	59.3						
gcpm5	F: TCCTGAACCACTAGACGATGG <sup>a</sup>	53.3	457	457	457	465	469	444
	R: TTTCGATAACTTCTTGATCCCTC T <sup>a</sup>	54.3						
gcpm6	F: CTCGCACTAAGCTCGGAAAT <sup>a</sup>	57.3	475	475	475	475	561	477
	R: ACCATGGCGTTACTCTACCG <sup>a</sup>	53.9						
gcpm7	F: TGGTGTGTTGAATCCACAAT <sup>a</sup>	53.2	297	297	289	322	322	322
	R: CCCCTTTCCAATAATATCCA <sup>a</sup>	55.9						
gcpm8	F: TCCAAAGAGGAATCCTTCCA <sup>a</sup>	53.3	237	237	237	313	237	237
	R: CCAGCTCTACTGGGGTTA <sup>a</sup>	55.3						
gcpm9	F: TCCAGGACTTCGAAAGGGTA <sup>a</sup>	53.4	492	492	492	492	486	511
	R: ACACGATACCAAGGCAAACC <sup>a</sup>	53.7						
gcpm10	F: CCGTGTTATCCGCTACATT <sup>a</sup>	54.1	227	227	228	257	225	228
	R: TCGTCTAAAATGCCTATACGAA CTC <sup>a</sup>	54.9						
gcpm11	F: CTGGACGATCCTATTTGGTCA <sup>b</sup>	53.7	220	220	220	238	220	202
	R: TCCAGCTCCGTATGAAGGTC <sup>b</sup>	53.8						
gcpm12	F: GCAGAATACCGTCACCCATT <sup>b</sup>	53.6	316	373	259	373	259	202
	R: ATTTTGTCCGGATCCTCCT <sup>b</sup>	53.8						
gcpm13	F: AAGATTTTATGAAGATACCGA AAAT <sup>a</sup>	52.5	484	484	466	461	456	465
	R: CGGTCAAATTCGAGGAAAGA <sup>a</sup>	55.3						
gcpm14	F: TGGTTAGTTTCACCGGATTCA <sup>b</sup>	54.2	208	208	208	208	208	178
	R: TTTTGTAGCCCATTTTAAGGA <sup>b</sup>	54.6						

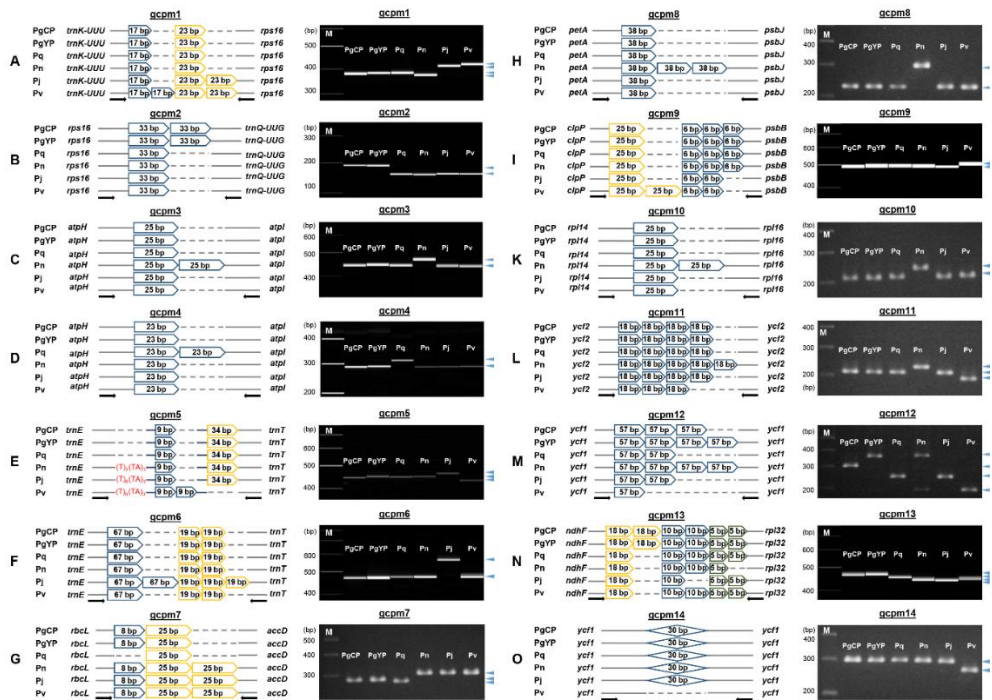
**Abbreviations:** PCR, polymerase chain reaction; PgCP, *P. ginseng* cv. CP; PgYP, *P. ginseng* cv. YP; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*.

<sup>a</sup>Primers developed in our previous study<sup>17</sup>, and <sup>b</sup>primers developed in this study

**Table 1-4.** Marker combinations for each *Panax* species.

Marker	Positions		Allele types for each species					
	Regions	CDS/IGS	PgCP	PgYP	Pq	Pn	Pj	Pv
gcpm1	<i>trnK</i> -UUU- <i>rps16</i>	IGS	C	C	C	D	B	A
gcpm2	<i>rps16</i> - <i>trnQ</i> -UUG	IGS	A	A	B	B	B	B
gcpm3	<i>atpH</i> - <i>atpI</i>	IGS	B	B	B	A	B	B
gcpm4	<i>atpH</i> - <i>atpI</i>	IGS	B	B	A	B	B	B
gcpm5	<i>trnE</i> - <i>trnT</i>	IGS	B	B	B	B	A	C
gcpm6	<i>trnE</i> - <i>trnT</i>	IGS	B	B	B	B	A	B
gcpm7	<i>rbcL</i> - <i>accD</i>	IGS	B	B	C	A	A	A
gcpm8	<i>petA</i> - <i>psbJ</i>	IGS	B	B	B	A	B	B
gcpm9	<i>clpP</i> - <i>psbB</i>	IGS	B	B	B	B	B	A
gcpm10	<i>rpl14</i> - <i>rpl16</i>	IGS	B	B	B	A	B	B
gcpm11	<i>ycf2</i>	CDS	B	B	B	A	B	C
gcpm12	<i>ycf1</i>	CDS	B	A	C	AD	C	D
gcpm13	<i>ndhF</i> - <i>rpl32</i>	IGS	A	A	B	C	D	C
gcpm14	<i>ycf1</i>	CDS	A	A	A	A	A	B

**Abbreviations:** CDS, coding sequences; IGS, intergenic spaces; PgCP, *P. ginseng* cv. CP; PgYP, *P. ginseng* cv. YP; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*. A, B, C, and D refer to different allele types.



**Figure 1-4. Validation of 14 molecular markers derived from InDel regions of five *Panax* chloroplast genomes.**

Schematic diagrams indicate InDel regions between *Panax* species. Tandem repeats and inserted sequences are designated by pentagons and diamonds, respectively. Dotted and solid lines indicate deleted and conserved sequences; left and right black arrows indicate forward and reverse primers, respectively. The 14 InDel markers are denoted gcpm1 to gcpm14. Different alleles are shown via capillary electrophoresis (A – F, I, N), and agarose gel electrophoresis (G, H, K – M, O). Abbreviated species names shown on schematic diagrams and amplicons: PgCP, *P. ginseng* cv. CP; PgYP, *P. ginseng* cv. YP; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*; M, 100-bp DNA ladder.



## DISCUSSION

### Repetitive sequences in *Panax* chloroplast genomes

Repeat structure plays an important role in genomic rearrangement and divergence in chloroplast genomes via illegitimate recombination and slipped-strand mispairing (Asano *et al.* 2004; Kim and Lee 2004). SSRs are direct tandem repeated DNA sequences consisting of short (1–10 bp) nucleotide motifs. The highly polymorphic, non-recombinant and uniparentally inherited nature of SSRs means they are often used as genetic molecular markers for population genetics (Doorduyn *et al.* 2011; He *et al.* 2012), and the study of ecology and evolution. In this study, I found 207 SSRs (Table 1-1), varying in number and type between five major *Panax* species.

It is considered that species divergence in chloroplast genomes is mostly caused by repetitive sequences. Indeed, I found many repetitive sequences associated with polymorphic chloroplast sites among the five major *Panax* species studied. Among the many repeats in the *ycf3* gene, the most divergent target region was a 57-bp TR. TRs such as this can be used to develop molecular markers for the identification and authentication of *Panax* species (Kim *et al.* 2015a).

### Comparative analysis of *Panax* chloroplast genomes

Polymorphism at the intraspecies level polymorphism is very low compared to that at the interspecies level. In *P. ginseng*, 12 chloroplast genomes derived from different cultivating varieties revealed over 99.9% sequence similarity, and only 6 InDels and 6 SNPs (Kim *et al.* 2015a). Meanwhile, of a total of 201 InDels and 962 SNPs identified among five *Panax* chloroplast genome sequences, 34–124 InDels and 141–560 SNPs were shared between two *Panax* species (Table 1-2). However, the

chloroplast genome is highly conserved within *Panax* species, and has high similarity ( $\geq 98.9\%$ ) at the nucleotide sequence level.

In chloroplast genomes, nucleotide substitution has been used to examine plant evolution and genome differentiation between species (Kim *et al.* 2013; Wolfe *et al.* 1987). These InDel events are mainly attributed to the repetition of an adjacent sequence, probably resulting from slipped-strand mispairing in DNA replication (Leseberg and Duvall 2009). InDels play a major role in genome size evolution and are increasingly used in phylogenetic studies (Britten *et al.* 2003; Grover *et al.* 2008). S/I ratios have been reported to increase with divergence time between genomes (Chen *et al.* 2009). In this study, I identified many SNPs and InDels from the complete chloroplast genomes of five major *Panax* species; S/I ratios ranged between 4.06 (*P. ginseng* versus *P. quinquefolius*) and 5.64 (*P. quinquefolius* versus *P. japonicus*) (Table 1-2). Along with their similar morphologies and distributions, this indicates that *P. ginseng* is closely related to *P. quinquefolius* (S/I = 4.06), and is consistent with previous reports (Nguyen *et al.* 2015; Zhu *et al.* 2003). *P. quinquefolius* is highly divergent from *P. japonicus* compared to others (S/I = 5.64).

### **Use of molecular markers to authenticate ginseng species**

Biological diversity is a valuable and vulnerable natural resource. The first steps towards protecting and benefiting from national biodiversity are to sample, identify, and study biological specimens. The use of molecular markers (i.e., DNA barcoding) has a powerful role to play in attaining many of the Millennium Development Goals, and reaching the objectives of the Convention on Biological Diversity, by sustaining natural resources, protecting endangered species, and identifying pests and pathogens at any life stage so as to more easily control them. DNA barcoding can also help to

monitor food, water and environment quality by studying contaminating organisms.

Although relatively new, the use of molecular markers is well on its way to being accepted as a global standard for species identification, and will play a major role in the future of taxonomy (Zuo *et al.* 2011). DNA-based species identification offers enormous potential for the biological scientific community, educators, and the interested public. The complete chloroplast genome has a conserved sequence from 110–160 kbp, which far exceeds the length of commonly used molecular markers and provides greater variation to discriminate between closely related species (Li *et al.* 2015). The chloroplast genome is smaller in size and hundreds times of higher copy numbers in a cell compared with the nuclear genome and has solid interspecific divergence with lower intraspecific variation, which makes chloroplast genome-based marker is suitable as DNA barcoding target (Li *et al.* 2015).

It is important to be able to authenticate natural health products (NHPs) from legal, economic, health and conservation viewpoints. NHPs are often trusted by the public to be safe; however, adulterated, counterfeit and low quality products can seriously threaten consumer safety (Mine and Young 2009; Wallace *et al.* 2012). Ginseng plants have high economic and medicinal value, thus there are many EMA ginseng products. The use of molecular methods to accurately identify the origins of ginseng products is important to the development of the ginseng industry in those countries where ginseng is cultivated, such as Korea, the USA, China, Japan, and Vietnam. Recently, the chloroplast genome sequences *rbcL*, *matK*, *trnH-psbA*, and *trnL-trnF*, and a nuclear internal transcribed spacer (ITS) of nuclear 45S ribosomal RNA genes, have successfully been used as molecular markers for several plant species (Hollingsworth *et al.* 2009; Li *et al.* 2011),

but these loci have little variability in *Panax* (Zuo *et al.* 2011). Although many studies have sought to authenticate ginseng (Chen *et al.* 2013; Choi *et al.* 2011; Kim *et al.* 2012; Ma *et al.* 2007), application remains limited because of a lack of genome information and comprehensive comparative genomic analysis against related *Panax* species.

In this study, I developed chloroplast-derived, species-specific markers for each of five major *Panax* species. Among 14 molecular markers, three markers were specific to *P. notoginseng*, and two were specific to *P. vietnamensis*. *P. ginseng*, *P. quinquefolius* and *P. japonicus* could also each be identified by species-specific markers (Table 1-4; Fig. 1-4). I discovered different 57-bp tandem repeats in the *ycf1* gene of different *Panax* chloroplast genomes, and this polymorphism proved powerful in the identification of *Panax* species (Table 1-4; Fig. 1-4M). However, the *gcpm12* marker, derived from the *ycf1* gene, unexpectedly produced two amplified bands from *P. notoginseng*, of which the A allele was the expected allele of the *P. ginseng* type; allele D was not expected, but was same as the allele of *P. Vietnamese*, although all other species gave rise to a single band (Table 1-4; Fig. 1-4A, M). I assume that these unexpected bands are derived from heteroplasmy of the target sequences caused by different IRA and IRB regions related to *ycf1* gene sequences; alternatively, they could be caused by chloroplast DNAs transferred into the nuclear or mitochondrial genomes (Goremykin *et al.* 2008; Massouh *et al.* 2016; Matsuo *et al.* 2005; Park *et al.* 2014). In our previous study (Kim *et al.* 2015a), we also found intra-species polymorphism in *P. ginseng* at this position, indicating that, although highly informative, the marker designed from this region may be confusing for species authentication. Overall, I suggest that intra-species polymorphism and a combination of several markers should be considered for credible authentication between different species.



## REFERENCES

- Allen G, Flores-Vergara M, Krasynanski S, Kumar S, Thompson W (2006)  
A modified protocol for rapid DNA isolation from plant tissues using  
cetyltrimethylammonium bromide. *Nature Protocols* 1:2320-2325
- Artyukova E, Kozyrenko M, Reunova G, Muzarok T, Zhuravlev YN (2000)  
RAPD analysis of genome variability of planted ginseng, *Panax ginseng*.  
*Molecular Biology* 34:297-302
- Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki KI (2004)  
Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*)  
chloroplast genome: a comparative analysis of four monocot chloroplast  
genomes. *DNA Research* 11:93-99
- Britten RJ, Rowen L, Williams J, Cameron RA (2003) Majority of  
divergence between closely related DNA samples is due to indels.  
*Proceedings of the National Academy of Sciences* 100:4661-4665
- Chan T, But P, Cheng S, Kwok I, Lau F, Xu H (2000) Differentiation and  
authentication of *Panax ginseng*, *Panax quinquefolius*, and ginseng  
products by using HPLC/MS. *Analytical Chemistry* 72:1281-1287
- Chen JQ, Wu Y, Yang H, Bergelson J, Kreitman M, Tian D (2009) Variation  
in the ratio of nucleotide substitution and indel rates across genomes in  
mammals and bacteria. *Molecular Biology and Evolution* 26:1523-1531
- Chen X, Liao B, Song J, Pang X, Han J, Chen S (2013) A fast SNP  
identification and analysis of intraspecific variation in the medicinal  
*Panax* species based on DNA barcoding. *Gene* 530:39-43
- Choi HI, Kim NH, Kim JH, Choi BS, Ahn IO, Lee JS, Yang TJ (2011)  
Development of reproducible EST-derived SSR markers and assessment  
of genetic diversity in *Panax ginseng* cultivars and related species.  
*Journal of Ginseng Research* 35:399

- Dong W, Liu J, Yu J, Wang L, Zhou S (2012) Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PloS One* 7:e35071
- Doorduyn L, Gravendeel B, Lammers Y, Ariyurek Y, Chin-A-Woeng T, Vrieling K (2011) The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Research* 18:93-105
- Elansary HO, Ashfaq M, Ali HM, Yessoufou K (2017) The first initiative of DNA barcoding of ornamental plants from Egypt and potential applications in horticulture industry. *PloS One* 12:e0172170
- Goremykin VV, Salamini F, Velasco R, Viola R (2008) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26:99-110
- Grover CE, Yu Y, Wing RA, Paterson AH, Wendel JF (2008) A phylogenetic analysis of indel dynamics in the cotton genus. *Molecular Biology and Evolution* 25:1415-1428
- He S, Wang Y, Volis S, Li D, Yi T (2012) Genetic diversity and population structure: implications for conservation of wild soybean (*Glycine soja* Sieb. et Zucc) based on nuclear and chloroplast microsatellite variation. *International Journal of Molecular Sciences* 13:12608-12628
- Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106:12794-12797
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput

- sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biology* 14:151
- Jung J, Kim KH, Yang K, Bang KH, Yang TJ (2014) Practical application of DNA markers for high-throughput authentication of *Panax ginseng* and *Panax quinquefolius* from commercial ginseng products. *Journal of ginseng research* 38:123-129
- Kim JH, Jung JY, Choi HI, Kim NH, Park JY, Lee Y, Yang TJ (2013) Diversity and evolution of major *Panax* species revealed by scanning the entire chloroplast intergenic spacer sequences. *Genetic Resources and Crop Evolution* 60:413-425
- Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* 11:247-261
- Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, Park HS, Yang TJ (2015a) Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS One* 10: e0117159
- Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS, Koh HJ, Waminal NE, Choi HI, Kim NH et al (2015b) Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific Reports* 5:15655
- Kim K, Nguyen VB, Dong JZ, Wang Y, Park JY, Lee SC, Yang TJ (2017) Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. *Scientific Reports* 7:4917
- Kim NH, Choi HI, Ahn IO, Yang TJ (2012) EST-SSR marker sets for practical authentication of all nine registered ginseng cultivars in Korea. *Journal of Ginseng Research* 36:298



- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research* 29:4633-4642
- Leseberg CH, Duvall MR (2009) The complete chloroplast genome of *Coix lacryma-jobi* and a comparative molecular evolutionary analysis of plastomes in cereals. *Journal of Molecular Evolution* 69:311-318
- Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences* 108:19641-19646
- Li R, Ma PF, Wen J, Yi TS (2013) Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. *PloS One* 8:e78568
- Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. *Biological Reviews* 90:157-166
- Ma KH, Dixit A, Kim YC, Lee DY, Kim TS, Cho EG, Park YJ (2007) Development and characterization of new microsatellite markers for ginseng (*Panax ginseng* CA Meyer). *Conservation Genetics* 8:1507-1509
- Massouh A, Schubert J, Yaneva-Roder L, Ulbricht-Jones ES, Zupok A, Johnson MT, Wright S, Pellizzer T, Sobanski J, Bock R (2016) Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of *Oenothera*. *The Plant Cell* 28:911-929
- Matsuo M, Ito Y, Yamauchi R, Obokata J (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast–nuclear DNA flux. *The Plant Cell* 17:665-675

- Mine Y, Young D (2009) Regulation of natural health products in Canada. *Food Science and Technology Research* 15:459-468
- Ngan F, Shaw P, But P, Wang J (1999) Molecular authentication of *Panax* species. *Phytochemistry* 50:787-791
- Nguyen B, Kim K, Kim YC, Lee SC, Shin JE, Lee J, Kim NH, Jang W, Choi HI, Yang TJ (2017) The complete chloroplast genome sequence of *Panax vietnamensis* Ha et Grushv (Araliaceae). *Mitochondrial DNA* 28:85-86
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PloS One* 7:e36869
- Park S, Ruhlman TA, Sabir JS, Mutwakil MH, Baeshen MN, Sabir MJ, Baeshen NA, Jansen RK (2014) Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics* 15:405
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106:411-422
- Wallace LJ, Boilard SM, Eagle SH, Spall JL, Shokralla S, Hajibabaei M (2012) DNA barcodes for everyday life: Routine authentication of Natural Health Products. *Food Research International* 49:446-452

- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences* 84:9054-9058
- Xu D, Abe J, Gai J, Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theoretical and Applied Genetics* 105:645-653
- Yamasaki K (2000) Bioactive saponins in Vietnamese ginseng, *Panax vietnamensis*. *Pharmaceutical Biology* 38:16-24
- Zhu S, Fushimi H, Cai S, Chen H, Komatsu K (2003) A new variety of the genus *Panax* from southern Yunnan, China and its nucleotide sequences of 18S ribosomal RNA gene and *matK* gene. *Journal of Japanese Botany* 78:86-94
- Zhu S, Zou K, Fushimi H, Cai S, Komatsu K (2004) Comparative study on triterpene saponins of Ginseng drugs. *Planta Medica* 70:666-677
- Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou S (2011) DNA barcoding of *Panax* species. *Planta Medica* 77:182

## **CHAPTER II**

**Comprehensive chloroplast genomes analysis of seven  
*Panax* species and development of an authentication  
system based on species-unique SNP markers derived  
from chloroplast genomes**

## ABSTRACT

*Panax* species are important herbal medicinal plants in the Araliaceae family. However, their evolutionary history, taxonomical relationships, and origins remain largely unresolved. Recently, we reported the complete chloroplast genomes and 45S nuclear ribosomal DNA (nrDNA) sequences from five *Panax* species, one (*P. quinquefolius*) from North America and four (*P. ginseng*, *P. notoginseng*, *P. japonicus*, and *P. vietnamensis*) from Asia. Here, I determined the chloroplast genome sequences of two more basal *Panax* species (*P. stipuleanatus* and *P. trifolius*). Phylogenetic analysis based on the nineteen complete chloroplast genomes revealed that nineteen Araliaceae species could be divided into two monophyletic lineages: one *Aralia-Panax* group and another group containing the seven remaining genera. I identified 1,128 SNPs in coding sequences (CDSs), distributed among 72 of the 79 protein-coding genes in the genomes of the seven *Panax* species. The other seven genes (including *psaJ*, *psbN*, *rpl23*, *psbF*, *psbL*, *rps18*, and *rps7*) were identical among the *Panax* species. I also discovered that twelve large chloroplast genome fragments were transferred into the mitochondrial genome, based on sharing over than 90% sequence similarity. The total size of transferred fragments was 60,331 bp, corresponding to approximately 38.6% of chloroplast genome. I developed 18 SNP markers from the chloroplast CDS regions that were not similar to regions in the mitochondrial genome. These markers included two or three species-specific markers for each species and can be used to authenticate all the seven *Panax* species from the others, thereby furthering efforts to protect the ginseng industry from economically motivated adulteration.

**Keywords:** Chloroplast genome, dCAPS markers, Ginseng authentication, *Panax* species, Phylogenetics.

## INTRODUCTION

*Panax* (ginseng) species are widely distributed from high altitude freeze-free regions including the Eastern Himalayas, the Hoang Lien Son, and the Annamite mountain range to the freezing winter regions of Northeastern Asia and North America. Ginseng contains many important pharmaceuticals that have been used in traditional medicine for thousands of years. Ginseng is also becoming one of the most important national agricultural commodities not only in Asian countries such as Korea, China, and Vietnam, but also in Russia, Canada, and the USA. Of the 14 known species in the *Panax* genus, five species, *Panax ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, and *P. vietnamensis*, have been utilized as expensive herbal medicines in Korea, the USA, China, Japan, and Vietnam. However, other species such as *P. stipuleanatus*, and *P. trifolius* still have limited information.

Notable therapeutic effects of ginseng on life-threatening diseases such as neurodegenerative (Cho 2012; Radad *et al.* 2006), and cardiovascular diseases (Zheng *et al.* 2012), diabetes (Xie *et al.* 2005), and cancer (Jung *et al.* 2012; Wong *et al.* 2015) have been well documented. Owing to the high pharmacological and economical value of ginseng, many economically motivated adulterations (EMAs) of ginseng products have been developed (Nguyen *et al.* 2017). Traditional methods for authentication of herb plants mainly depended on morphological and histological characteristics. However, morphological and histological authentication are not precise enough because of the similar morphological appearances between species or the different morphological character of intraspecies related to growing conditions. Moreover, almost all commercial ginseng products are sold in various forms such as dried root, powder, liquid extracts or other processed products, which are impossible to authenticate based on morphology.

Authentication based on ginsenoside profiling methods have been developed (Chan *et al.* 2000; Wang *et al.* 1999; Yang *et al.* 2016; Yuk *et al.* 2013). However, these applications are limited because ginsenosides are secondary metabolite and the accumulation of ginsenosides vary in quantity between different tissues (such as roots, leaves, stems, flower buds and berries) (Oh *et al.* 2014; Shi *et al.* 2007), cultivars (Lee *et al.* 2017), age (Shi *et al.* 2007; Xiao *et al.* 2015), and environmental conditions (Jiang *et al.* 2016; Kim *et al.* 2014b). Additionally, storage conditions and manufacturing processes also affect secondary metabolite accumulation in ginseng (Nguyen *et al.* 2017).

Chloroplasts are multifunctional organelles containing their own genetic material required for photosynthesis and carbon fixation. Chloroplast genomes are highly conserved in plants, with the quadripartite structure comprising two copies of inverted repeat regions (IRs) that separate the large and small single-copy (LSC and SSC) regions. The chloroplast genome size in angiosperm usually ranges from 115 to 165 kb (Leebens-Mack *et al.* 2005). Since the emergence of next-generation sequencing, the number of completely sequenced chloroplast genomes rapidly increased. As of September 2017, more than 1541 complete chloroplast genomes of land plants are available in the GenBank Organelle Genome Resources. Of these, five chloroplast genomes from *Panax* genus have been sequenced (Kim *et al.* 2017).

Sequence-based DNA markers have shown the advantages and power in species identification with high accuracy, simplicity, and time- and cost-efficiency (Nguyen *et al.* 2017). Various DNA markers have been applied for authentication of *Panax* species including nuclear genome-derived random amplified polymorphic DNA (RAPD) (Artyukova *et al.* 2000), microsatellite (Ma *et al.* 2007), and expressed sequence tag-simple sequence repeat (EST-SSR) (Choi *et al.* 2011; Kim *et al.* 2012) markers. However

these nuclear genome-derived DNA markers are usually utilized for intra-species level diversity. DNA markers based on the chloroplast genome have been widely used and considered as the best barcoding target for plant species identification (Li *et al.* 2015) because of their highly conserved structure and high copy numbers for easy detection. Chloroplast genome divergence is lower at intraspecific level and higher at interspecific level. Recently, chloroplast-derived DNA markers have been developed to authenticate ginseng, including single nucleotide polymorphism (SNP) and insertion or deletion (InDel) markers (Chen *et al.* 2013; Jung *et al.* 2014; Kim *et al.* 2015a; Nguyen *et al.* 2017). However, these markers are still of limited use due to the lack of genomic information for intra- and inter-species variations.

Recently, we obtained complete chloroplast genome and nuclear ribosomal DNA (nrDNA) for five major *Panax* species (Kim *et al.* 2015a; Kim *et al.* 2017) and two basal *Panax* species in Asia and in North America by *de novo* assembly using low-coverage whole-genome shotgun next-generation sequencing (dnaLCW) (Kim *et al.* 2015b). Using this information, we previously developed InDel-based authentication markers among the five species (Nguyen *et al.* 2017). Although these markers are easy to apply, their usefulness is somewhat limited by the relatively rich intra-species polymorphism at the InDel regions. In this study, we conducted a comprehensive comparative genomics study of the chloroplast genomes from the seven *Panax* species and identified 18 chloroplast coding sequence (CDS)-derived SNP markers that can be used to authenticate each of the seven species. This study provides valuable genetic information as well as a practical marker system for authentication of each *Panax* species that will be very helpful for regulating the ginseng industry.



## MATERIAL AND METHODS

### Plant materials and genomic DNA extraction

*P. ginseng* cultivars and *P. quinquefolius* plants were collected from the ginseng farm at Seoul National University in Suwon, Korea. *P. notoginseng* and *P. japonicus* plants were collected from Dafang County, Guizhou Province, and Enshi County, Hubei Province, China, respectively. *P. vietnamensis* and *P. stipuleanatus* plants were collected from Kon Tum and Lao Cai Province, Vietnam, respectively. *P. trifolius* plants were collected from North Eastern America. Leaves and roots were used for DNA extraction using a modified cetyltrimethylammonium bromide (CTAB) method (Allen *et al.* 2006). The quality and quantity of extracted genomic DNA was measured using a UV-spectrophotometer, and agarose-gel electrophoresis.

### Phylogenetic analysis

Phylogenetic tree construction and the reliability assessment of internal branches were conducted by the maximum likelihood (ML) method with 1,000 bootstrap replicates using MEGA6.0 (Tamura *et al.* 2013).

### Comparative analysis of 79 CDS genes between seven *Panax* species

The chloroplast genome sequences of 11 *P. ginseng* cultivars (ChP\_KM088019, YP\_KM088020, GU\_KM067388, GO\_KM067387, SP\_KM067391, SO\_KM067390, SU\_KM067392, SH\_KM067393, CS\_KM067386, HS\_KM067394, JK\_KM067389), 2 *P. quinquefolius* (KM088018, KT028714), 4 *P. notoginseng* (KP036468, KT001509, NC\_026447, KR021381), 1 *P. japonicus* (KP036469), 2 *P. vietnamensis* (KP036471, KP036470), 1 *P. stipuleanatus* (KX247147), and 1 *P. trifolius* (MF100782) were obtained from our previous studies (Kim *et al.* 2015a; Kim *et al.* 2017) and Genbank. Chloroplast CDS gene sequences were manually curated and extracted using Artemis (Rutherford *et al.* 2000). The

chloroplast CDS sequences were concatenated and aligned using MAFFT program (<http://mafft.cbrc.jp/alignment/server/>). The SNP of 79 CDS gene sequences were calculated by MEGA 6 (Tamura *et al.* 2013), then SNPs were located on the chloroplast CDS sequence maps of 7 *Panax* species using Circos v.0.67 (Krzywinski *et al.* 2009).

### **Identification of chloroplast gene insertion in mitochondria**

Mitochondrial genome of *Panax ginseng* were downloaded from NCBI (KF735063) and masked to chloroplast genome to eliminate BLAST hits of transferred genes between chloroplast and mitochondrial genomes. The map of *Panax* species chloroplast and mitochondrial genomes and the fragment of gene transfers were drawn using Circos v.0.67 (Krzywinski *et al.* 2009).

### **Development and validation of dCAPS markers**

To authenticate between seven *Panax* species as interspecies polymorphism using chloroplast CDS sequences, dCAPS primers were designed based on SNP polymorphic sites after eliminating intraspecies polymorphic sites, and chloroplast gene transfer regions. The dCAPS were designed to create a suitable restriction enzyme cutting sites using the dCAPS Finder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>), and the specific primers were designed using the Primer3 program (<http://bioinfo.ut.ee/primer3-0.4.0/>).

The polymerase chain reaction (PCR) was carried out in a 25 µl reaction mixture containing 2.5 µl of 10x reaction buffer, 1.25 mM deoxynucleotide triphosphate (dNTP), 5 pmol of each primer, 1.25 units of *Taq* DNA polymerase (Inclone, Korea), and 20 ng of DNA template. The PCR reaction was performed in thermocyclers using the following cycling parameters: 94°C (5 min); 35 cycles of 94°C (30 s), 56 – 62°C (30 s); 72°C (30 s), then 72°C (7 min). PCR products were visualized on agarose gels (2.0 – 3.0%) containing safe gel stain (Inclone, Korea).

An analytical restriction enzyme reaction were performed in a volume of 10  $\mu$ l containing 5  $\mu$ l of PCR product, 1  $\mu$ l of 10x restriction enzyme buffer, 0.3  $\mu$ l of 10 unit restriction enzyme. The reaction mixture were incubated at the optimum temperature for 3 hours or overnight, then visualized on agarose gels (2.0 – 3.0%) containing safe gel stain.

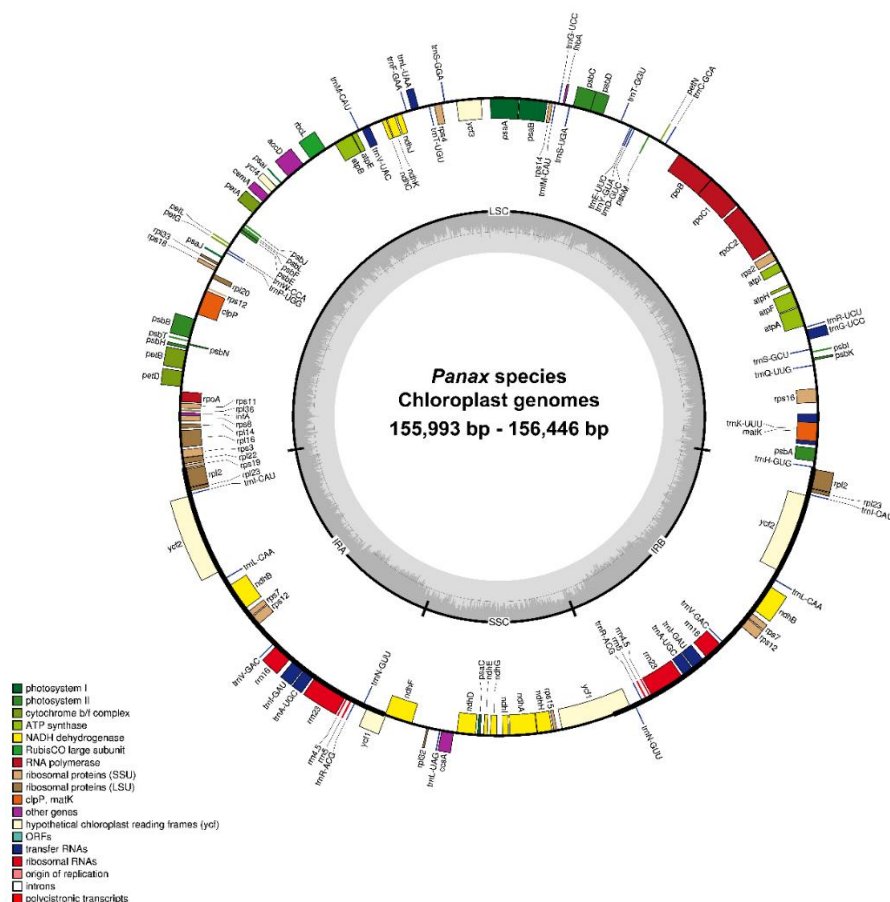
## RESULTS

### **Characteristics of the complete chloroplast genomes of seven *Panax* species**

Complete chloroplast genome length of seven *Panax* species ranged from 155,993 bp to 156,466 bp (Table 2-1). The chloroplast genome of seven *Panax* species showed a typical quadripartite structure, consisting of a pair of IRs separated by the LSC and SSC regions (Fig. 2-1). There is no structural variations except small InDels and SNPs. It contains 113 functional genes, including 79 protein-coding genes, 30 transfer RNA (tRNA) genes and 4 ribosomal RNA (rRNA) genes, and the gene map for seven *Panax* chloroplast genomes were shown as Fig. 2-1.

**Table 2-1.** Chloroplast genome sequences used in this study.

Species	Whole genome NGS sequence amounts (Mb)	Sequence reads used		Chloroplast genome length (bp)
		Amounts (Mb)	Chloroplast Coverage (x)	
<i>P. ginseng</i>	10,418	505	97	156,248 (KM088019)
<i>P. quinquefolius</i>	3,557	1,010	127	156,088 (KM088018)
<i>P. notoginseng</i>	5,619	2,811	246	156,466 (KP036468)
<i>P. japonicus</i>	5,738	2,870	237	156,188 (KP036469)
<i>P. vietnamensis</i>	7,541	4,586	1,005	155,993 (KP036470)
<i>P. stipuleanatus</i>	2,218	599	154	156,064 (KX247147)
<i>P. trifolius</i>	14,657	2,300	993	156,157 (MF100782)

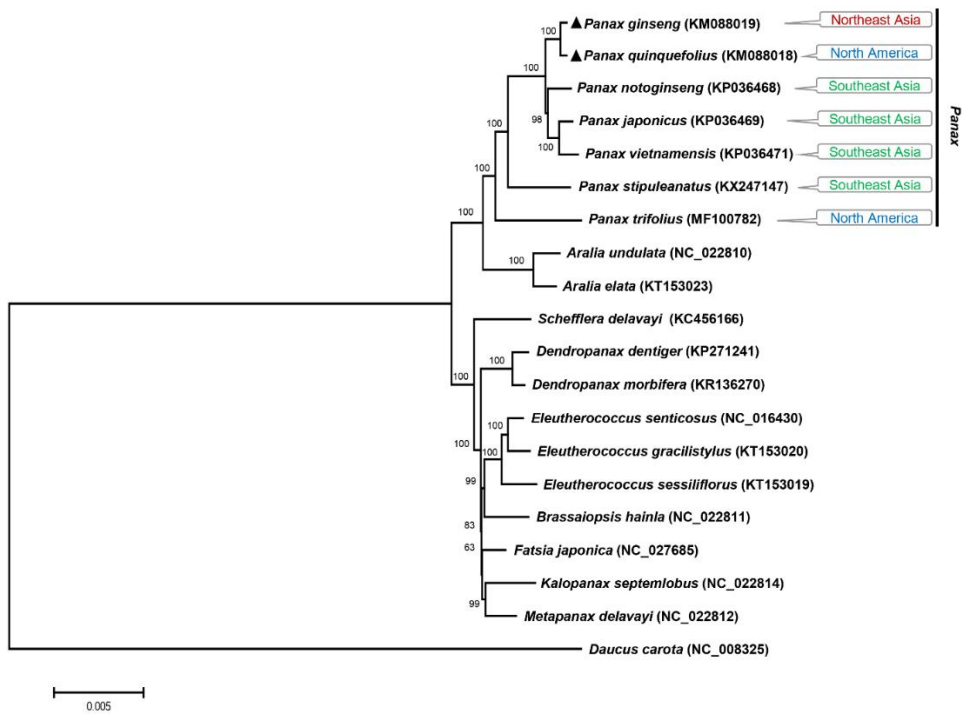


**Figure 2-1. Complete chloroplast genomes of seven *Panax* species.**

Colored boxes show conserved chloroplast genes, classified based on product function. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. Dashed area in the inner circle indicates the GC content of the chloroplast genome.

### **Phylogenomic analysis for 19 complete chloroplast genomes of Araliaceae species**

Phylogenetic relationship was inferred using entire chloroplast genome sequences from 19 species in the Araliaceae family indicated that nine genera in Araliaceae were divided into two typical monophyletic lineages consisting of the *Aralia-Panax* group and the other group with seven remaining genera (Fig. 2-2). Species of each genus, *Panax*, *Aralia*, *Schefflera*, *Dendropanax*, *Eleutherococcus*, *Brassaiopsis*, *Fatsia*, *Kalopanax* and *Metapanax* were grouped accordingly. Based on the phylogenetic tree, the seven *Panax* species were divided into some subgroups, in which *P. stipuleanatus* and *P. trifolius* were divergent from common ancestor earlier than other five *Panax* species (Fig. 2-2).



**Figure 2-2. ML phylogenetic tree of *Panax* and *Panax*-related species in the Araliaceae family based on entire chloroplast genome sequences.** Numbers in the nodes are the bootstrap support values from 1000 replicates. Black triangles indicate tetraploid *Panax* species. The chloroplast sequence of carrot (*Daucus carota*) was used as an outgroup.



### **SNPs in chloroplast genomes of seven *Panax* species**

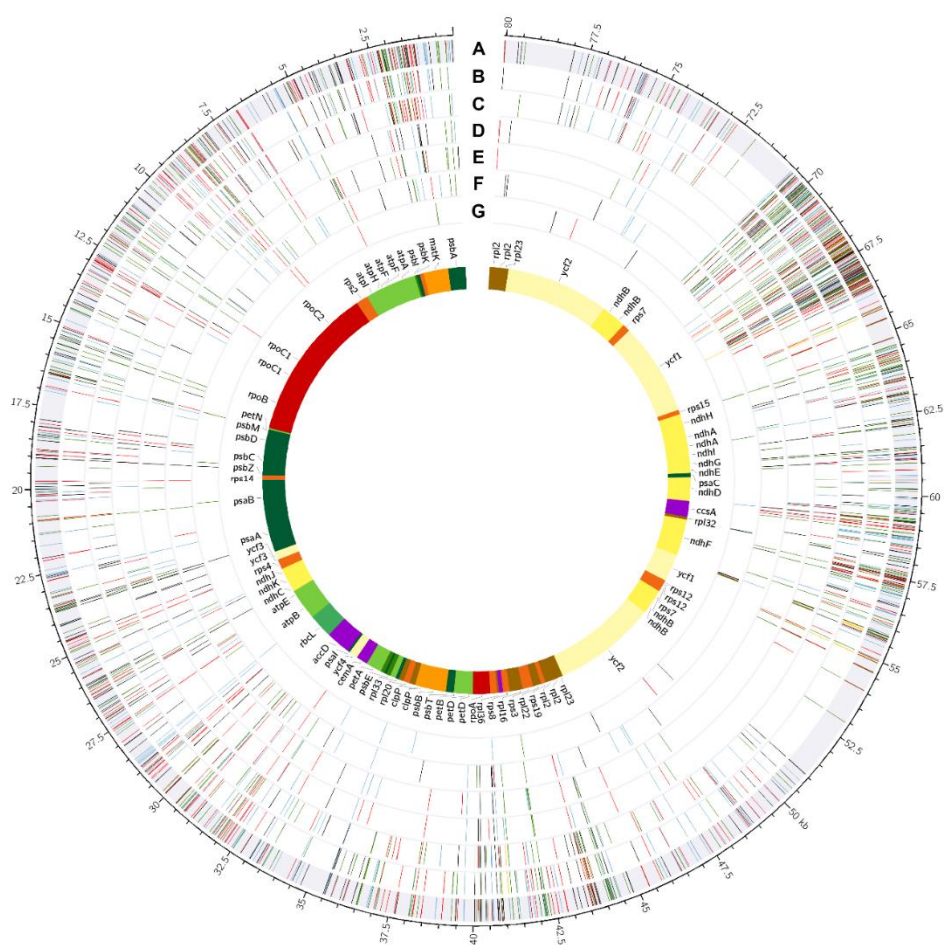
Single nucleotide polymorphism (SNP) were calculated in chloroplast genomes for 7 *Panax* species to compare the polymorphism between species and develop SNP-derived markers for authentication. A total of 1,783 SNP sites were identified in whole chloroplast genome sequences and 1,128 SNP sites were identified in CDS gene sequences. Even though numbers of SNPs in CDS gene regions were lower than 50% of those in whole chloroplast genome sequences of each species (Table 2-2). The two closest tetraploid species (*P. ginseng* and *P. quinquefolius*) had a lower number of SNPs in both CDSs and whole chloroplast genome sequences than any other pair (Table 2-2). *P. trifolius* had the highest numbers of SNPs in both CDS and whole chloroplast sequences in comparison with each of the six other species (Table 2-2). SNP numbers were distributed in the 71 of the 79 CDS gene sequences of seven *Panax* species except 8 highly conserved genes including *petN*, *psaJ*, *psbN*, *rpl23*, *psbF*, *psbL*, *rps18*, and *rps7* (Fig. 2-3). SNP density were lower in IR regions than in LSC and SSC regions (Fig. 2-3).

**Table 2-2.** Numbers of SNPs among seven *Panax* chloroplast genomes.

	<i>P. ginseng</i>	<i>P. quinquefolius</i>	<i>P. notoginseng</i>	<i>P. japonicus</i>	<i>P. vietnamensis</i>	<i>P. stipuleanatus</i>	<i>P. trifolius</i>
<i>P. ginseng</i>	/	131	460	495	531	1157	1485
<i>P. quinquefolius</i>	59	/	493	496	518	1145	1479
<i>P. notoginseng</i>	171	210	/	476	535	1159	1514
<i>P. japonicus</i>	183	220	183	/	316	1150	1513
<i>P. vietnamensis</i>	246	245	243	157	/	1196	1555
<i>P. stipuleanatus</i>	497	534	522	524	566	/	1484
<i>P. trifolius</i>	594	610	621	624	664	639	/

Total: 1783 SNP sites in the whole chloroplast genomes and 1128 in CDS regions of seven species

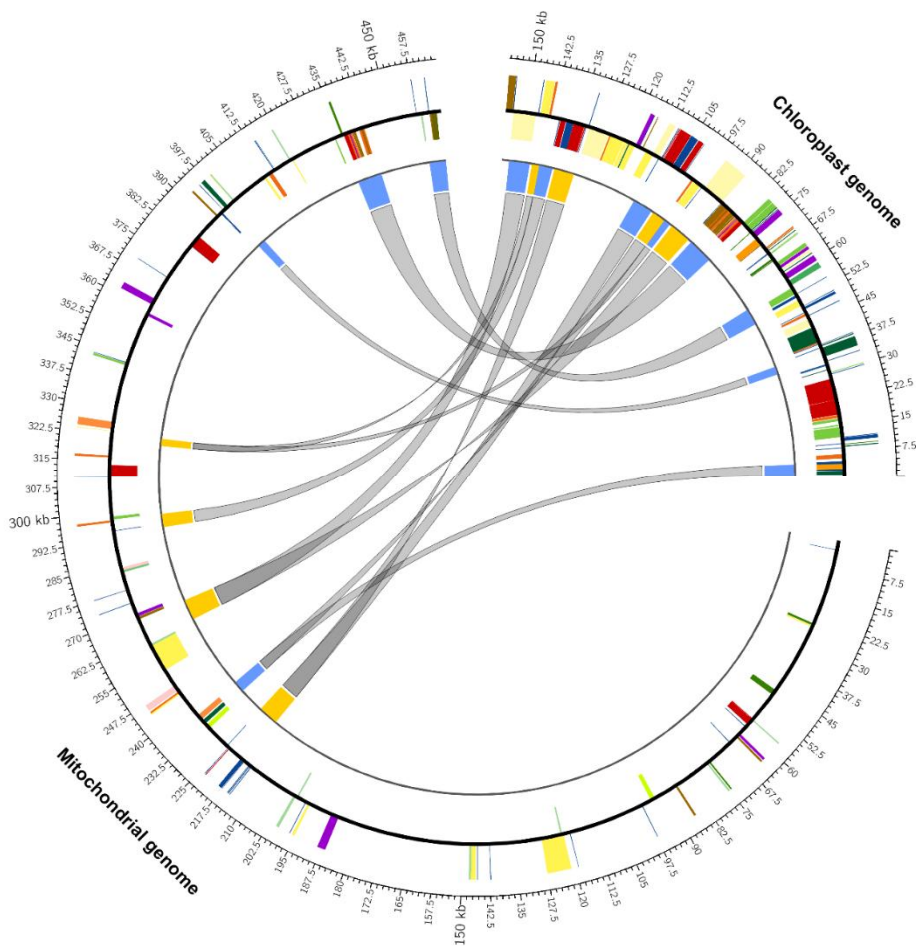
The upper triangle shows the total number of SNPs in whole chloroplast genome, and the lower triangle indicates the number of SNPs in the 79 CDS regions of each species.



**Figure 2-3. Single nucleotide polymorphic sites in 79 CDS genes of 7 *Panax* species.** The inner track showing 79 chloroplast CDS genes, track A represents total SNPs in all 7 *Panax* species. The track B – G represent SNPs in *P. trifolius*, *P. stipuleanatus*, *P. vietnamensis*, *P. japonicas*, *P. notoginseng*, and *P. quinquefolius* compared to *P. ginseng*, respectively. Red, green, blue and black lines on each track indicate the four kinds of SNPs (T, A, C and G nucleotides), respectively. Yellow lines indicate InDel regions.

### **Characterization of chloroplast genome transfer into mitochondrial genome**

Mitochondrial genome sequence of *P. ginseng* retrieved from GenBank is 464,680 bp with approximately 3 times larger than chloroplast genome and consists of 94 functional genes (Fig. 2-4). I identified twelve large chloroplast genomes fragments in mitochondrial genome. The fragments ranged from 2,297 to 8,250 bp with  $\geq 90\%$  sequence identity with the original chloroplast counterparts (Fig. 2-4). The total size were 60,331 bp, which correspond approximately 38.6% of chloroplast genome (Fig. 2-4). IR regions showed higher numbers of gene transfer to mitochondrial genome than those of SSC and LSC. The gene transfer regions covered almost 49 chloroplast genes and intergenic regions (Fig. 2-4).



**Figure 2-4. Schematic representation of gene transfer between chloroplast and mitochondrial genome of *Panax* species.** Each gray line within the circle shows the regions of chloroplast genome that have been inserted into different locations of mitochondrial genome. Colored boxes show conserved chloroplast genes, classified based on product function. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise.

### **Identification of species-specific SNP markers for authentication of seven *Panax* species**

A total of 18 dCAPS markers were developed from species-specific SNP targets among 7 *Panax* species. Each SNP target is derived from coding CDS sequence and show unique polymorphism in one species among seven *Panax* species. At least two dCAPS markers were developed for each species (Table 2-3). Each of these 18 markers showed the expected band sizes before and after restriction enzyme digestion (Fig. 2-5). Markers Pgdm1 – 3 that were derived from the *rpl20*, *ndhK*, and *rps15* gene sequences, respectively, were specific to *P. ginseng* and resulted in different band sizes when digested compared to other species (Fig. 2-5). Markers Pqdm4 – 6 were derived from *rpoC1*, *ndhA*, and *ndhK* sequences, respectively, and resulted in a unique digestion pattern for *P. quinquefolius* (Fig. 2-5). Markers Pndm7 – 9 were derived from *rpoC1*, *rpoC2*, and *ndhK* sequences, respectively, and resulted in a unique digestion pattern for *P. notoginseng* (Fig. 2-5). Markers Pjdm10 and 11 were derived from the *rpoC2* and *rpoB* sequences, respectively, and their digestion pattern was unique for *P. japonicus*, while markers Pvdm12 and 13 were derived from *rpoC2* and *ndhH* genes, respectively, and resulted in a digestion pattern that was unique for *P. vietnamensis* (Fig. 2-5). Markers Psdm14 – 16 were derived from *psbB*, *rpoC1*, and *rpoB*, respectively, and resulted in a digestion pattern that was unique for *P. stipuleanatus* (Fig. 2-5). Two markers, Ptdm17 and 18, were derived from *ndhA* and *rpoC1*, respectively, and resulted in a unique digestion pattern for *P. trifolius* (Fig. 2-5). All 18 markers were practical and successful for distinguishing among the seven *Panax* species, and can therefore be applied to ginseng species authentication.

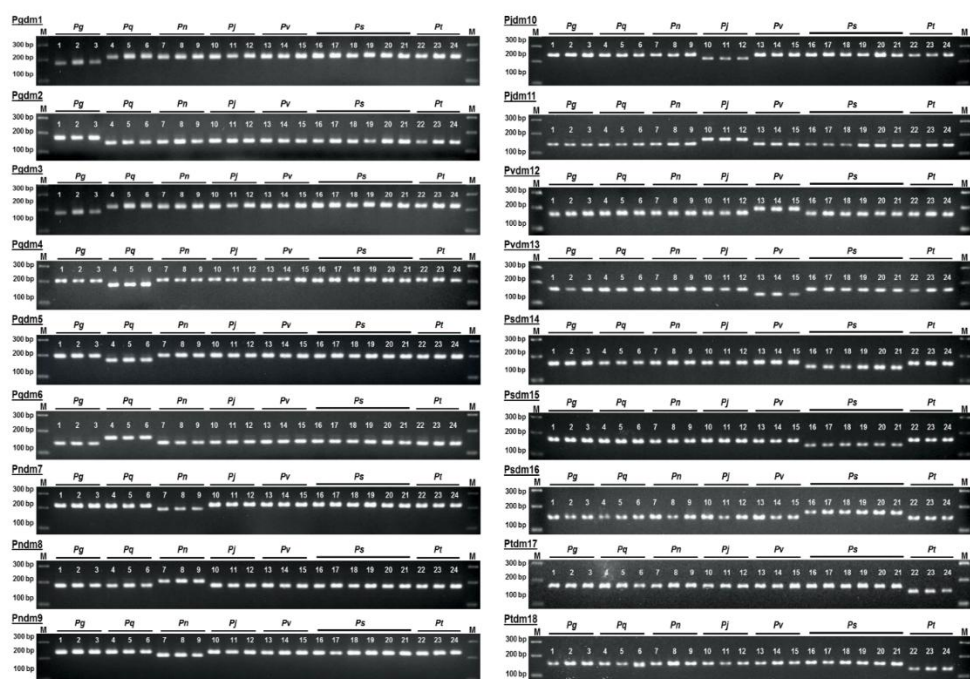
**Table 2-3.** Details for the dCAPS markers developed to authenticate *Panax* species.

Marker ID	Primer sequence (5'-3')	Location	T <sub>m</sub> (°C)	PCR product size (bp)	Digestion enzyme	Target SNP						
						Pg	Pq	Pn	Pj	Pv	Ps	Pt
Pgdm1	GTTTAAATTATTCCGGTGGATTCTT	<i>rpl20</i>	59.2	170	Cla1	A	G	G	G	G	G	G
	GTAGCCTATAGTTATAGTAGATTAATCGA		63.4									
Pgdm2	GTCCGCTTGTCTAGGACTCG	<i>ndhK</i>	62.5	177	Cla1	A	G	G	G	G	G	G
	CAAAATTCAGTTATTTCAACTACATCAAT		60.5									
Pgdm3	ATCCAACCGACCAATTAATTCCTTTA	<i>rps15</i>	59.2	219	Sma1	C	T	T	T	T	T	T
	TTGAAAGAGGAAAAACAAAGACACCC		62.5									
Pqdm4	TATGACCGTCCCTCATCGGTTGTCTG	<i>rpoC1</i>	69.1	212	Sal1	G	A	G	G	G	G	G
	CATTCAGATAGATGGGGGTAAACTA		62.5									
Pqdm5	CTCGTAAACCACCTAAAAAGGAAT	<i>ndhA</i>	60.1	206	Cla1	C	T	C	C	C	C	C
	TCGTTTATTCAAGTATCGGACCATCG		64.2									
Pqdm6	TTCCGGCTGTTAAAATTAGGTCAGC	<i>ndhK</i>	63	167	Alu1	T	C	T	T	T	T	T
	TCTTTCAAATTGGTCAAGACTCTCT		60.9									
Pndm7	CCTATTTACACAAATACCCCGTCGA	<i>rpoC1</i>	64.2	223	Sal1	T	T	C	T	T	T	T
	ATTAGTTCGTAAAGGATTCAATGCAG		61.6									
Pndm8	TTCATTTGATCTTGATCCTTG TG	<i>rpoC2</i>	57.5	216	Hind3	A	A	G	A	A	A	A
	TCCACTTTGAATTTTAAAGAGAAGCT		60.0									
Pndm9	ATCGACAGGAATTAGCTTATCGAC	<i>ndhK</i>	61.8	238	Cla1	G	G	A	G	G	G	G
	ACGATTCGACTTTGATCGTTATCGA		62.5									
Pjdm10	TGGATATCTCCAGAAAATATTTTAAGTAC	<i>rpoC2</i>	62.0	250	Sal1	A	A	A	T	G	A	A
	AGGATTTGATTGAGTATCGAGGAG		61.8									
Pjdm11	AGTCCGACATTTATTCCTTCAGAC	<i>rpoB</i>	61.8	172	Rsa1	T	T	T	C	T	T	T
	GTTTTGGATCGAACTAATCCATTGGT		63.2									
Pvdm12	TGCGCGAATCTCAGCAATCACTAG	<i>rpoC2</i>	65.3	195	Spe1	T	T	T	T	C	T	T
	AAATTCAATGAGGATTTGGTTCAT		56.7									

Pvdm13	CATAAGGTAAATACTGTATAATTGATCG TATGATAGTCAATCTGGGTCCTCA	<i>ndhH</i>	61.3 61.8	170	Cla1	G	G	G	G	A	G	G
Psdm14	AACCTTCTTTGGATTTGCCCAAGCT CACGCTGGATTACAGATTGTACT	<i>psbB</i>	64.2 61.8	166	Hind3	C	C	C	C	C	T	C
Psdm15	GAAGCCACAAAGGACTATCTAAATG GTCGGGGTATTTGTGTAAATAGGT	<i>rpoC1</i>	62.5 61.8	180	EcoR1	G	G	G	G	G	A	G
Psdm16	TAAGCTTCCTTCCTATTAATCTGG GAATT CATATTAGAGCTCGCCAGGAAGTA	<i>rpoB</i>	64.8 63.5	179	EcoR1	C	C	C	C	C	T	C
Ptdm17	TATGTACGGAATAGAAAGATTCCAAGC CGAGTGTGAGAGATTACCTTTTGA	<i>ndhA</i>	63.7 61.8	187	Alu1	C	C	C	C	C	C	T
Ptdm18	CGCTCTATTTAGCAATACGGGATA GCAATAGAGCTTTTCCAGACATTT	<i>rpoC1</i>	61.8 60.1	162	EcoR5	C	C	C	C	C	C	T

**Abbreviations:** Pg, *Panax ginseng*; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*; Ps, *P. stipuleanatus*; Pt, *P. trifolius*.





**Figure 2-5. Validation of 18 dCAPS markers derived from CDS SNP regions of seven *Panax* chloroplast genomes.** The 18 denoted dCAPS markers, Pgdm1 – 3, Pqdm4 – 6, Pndm7 – 9, Pjdm10 and 11, Pvdml2 and 13, Psdm14 – 16, Ptdm17 and 18 are unique specific for *P. ginseng*, *P. quinquefolius*, *P. notoginseng*, *P. japonicus*, *P. vietnamensis*, *P. stipuleanatus*, and *P. trifolius*, respectively. Abbreviated species names shown on amplicons: Pg, *P. ginseng*; Pq, *P. quinquefolius*; Pn, *P. notoginseng*; Pj, *P. japonicus*; Pv, *P. vietnamensis*; Ps, *P. stipuleanatus*; Pt, *P. trifolius*; M, 100-bp DNA ladder.

## DISCUSSION

### **Complete chloroplast genomes of seven *Panax* species derived from low-coverage whole-genome NGS data**

Chloroplast DNA sequences play an important role in genetic engineering (Cui *et al.* 2014), DNA barcoding (Dong *et al.* 2014), and in studying evolutionary relationships among plants (Huang *et al.* 2014; Li *et al.* 2013). More recently, with technical advances in DNA sequencing, the number of completely sequenced chloroplast genome has grown rapidly. However, the complete chloroplast genome sequences for many high value plant species is not available yet because of high cost (Cho *et al.* 2015). In our previous studies, we applied *de novo* assembly method using low-coverage whole-genome shotgun next-generation sequencing (dnaLCW) (Kim *et al.* 2015b) to obtain complete chloroplast genomes of seven *Panax* species. All seven chloroplast genome sequences were supported from average read-mapping coverage of 154x and 993x (Table 2-1). Chloroplast structure are identical with small size InDels and different numbers of SNPs in all seven *Panax* species (Kim *et al.* 2017). Complete chloroplast genomes of seven *Panax* species will provide more valuable genetic information for study of evolutionary relationship, breeding, and authentication of ginseng species.

### **Phylogenetic analysis**

The Araliaceae is a family of flowering plants comprising about 70 genera and approximately 750 species varying in habit from trees and shrubs to lianas and perennial herbs (Court 2006). Araliaceae speciation were suggested to be occurred in two particular regions of North America and South East Asia (Court 2006), and the diversification and speciation were associated with whole genome duplication (WGD) or polyploidy events (Freeling and Thomas 2006; Soltis *et al.* 2009; Wood *et al.* 2009). Previous studies indicated that two tetraploid *Panax* species, *P. ginseng* and *P.*

*quinquefolius*, have undergone two rounds of WGD (Choi *et al.* 2013; Kim *et al.* 2014a). The WGD event along with geographic and ecological isolation have together contributed to the diversification of *Panax* species (Shi *et al.* 2015). Traditionally, taxonomy of *Panax* based on the morphological characters has been controversial due to the complicated morphological variation between intra- and interspecies according to geographic and ecological environment.

My phylogenetic tree based on whole chloroplast genome clearly showed the evolutionary relationship between *Panax* species, and between genera in the Araliaceae family. In particular, my results indicated that the diploid species *P. trifolius*, which was divergent earlier from common ancestor and migrated to North America was not involved in the tetraploidization of *P. ginseng* and *P. quinquefolius*. The another diploid species *P. stipuleanatus*, which was earlier divergent from the five remaining species showed overlapped distribution with the three diploid species group in South East Asia including *P. notoginseng*, *P. vietnamensis* and *P. japonicus*. Two tetraploid species, *P. ginseng* and *P. quinquefolius*, which are involved in the recent second WGD were diverged from the three diploid species group and located in Northeastern Asia and North America due to geographic isolation (Fig. 2-2).

### **Comparative analysis of *Panax* chloroplast genomes**

SNP at the intraspecies level polymorphism is very low compared to that at the interspecies level. Numbers of SNP among 12 *P. ginseng* cultivar of whole chloroplast genomes were rare, and only 6 SNPs were identified in 12 *P. ginseng* cultivars (Kim *et al.* 2015a). Meanwhile, a total of 1,783 SNPs and 1,128 SNPs identified among seven *Panax* species at the whole chloroplast genome and CSD gene sequences, respectively (Table 2-2). Nevertheless, the chloroplast genomes are highly conserved with high

similarity ( $\geq 97.6\%$ ) at the nucleotide sequence level within *Panax* genus. In our previous study, we found that some chloroplast protein-coding genes are highly diverse or highly conserved among different Araliaceae species. Three genes, *infA*, *rpl22*, *rps19* and *ndhE* genes, were diverse showing large numbers of SNP polymorphism between different species. And some genes, including *atpF*, *atpE*, *ycf2*, and *rps15* genes, showed high non-synonymous mutation which might be related to evolution under positive selection (Kim *et al.* 2017). However, some genes showed highly conserved at the Araliaceae family level such as *petN*, *psaJ*, *psbN* and *rpl23* genes, or even at the Apiales order level such as *psbF* gene (Kim *et al.* 2017). The current study is in consistent with the previous study, and in addition to those five highly conserved genes, I found three more genes, *psbL*, *rps18* and *rps7* genes, which were highly conserved in 7 *Panax* species (Fig. 2-3).

### **Chloroplast genome fragments in mitochondrial genome**

The DNAs sequencing of different genomes (nuclear, chloroplast, and mitochondrial) has uncovered staggering amounts of intracellular gene transfer between them (Kleine *et al.* 2009; Timmis *et al.* 2004). Studies have shown that there is a high frequency of organelle DNA transfer to the nucleus in angiosperm (Hazkani-Covo *et al.* 2010; Park *et al.* 2014; Smith *et al.* 2011). Inter-organelle genome transfer from chloroplast to mitochondrial genomes is also reported recently as a common phenomenon in higher plants in the course of evolution (Gui *et al.* 2016; Park *et al.* 2014). I identified twelve large fragments of chloroplast genomes (representing 38.6% of chloroplast genome) in mitochondrial of *Panax* species, and these sequences covered a lot of gene and intergenic regions (Fig. 2-4). Genome transfer can cause the assembly errors of chloroplast or mitochondrial genomes due to the high similarity of gene sequences. Moreover, study of evolution or development of molecular markers from gene transferring regions can cause

confusing or bias results (Nguyen *et al.* 2017). To reduce this limitation, I examine all the gene transfer regions and eliminated all these regions before developing SNP-derived markers for authentication.

### **Use of dCAPS markers for ginseng species authentication**

DNA barcoding may be defined as use of short DNA sequences of nuclear or organelle genome for identification of species. DNA barcoding is a new technique and have been widely used as a biological tool in species identification, breeding, and evolution research (Zuo *et al.* 2011). Identification of plant species is important for standardizing food and herbal medicine, and help to prevent EMAs. Ginseng has high pharmacological and economical value, thus there are many potential of EMA ginseng products. Therefore, easily detectable, reliable and practical method to accurately identify the origins of ginseng products play an important role to the development of the ginseng industry.

The chloroplast genome is endemic to plant with smaller in size and hundreds times of higher copy numbers in a cell compared with the nuclear genome. Furthermore, chloroplast genome has high enough interspecific divergence and low intraspecific variation, thus chloroplast genome-based DNA barcodes are the best target for species authentication (Li *et al.* 2015). Recently, the chloroplast genome sequences have been used to develop marker for ginseng authentication (Chen *et al.* 2013; Jung *et al.* 2014; Ngan *et al.* 1999; Nguyen *et al.* 2017). However, the application remained in certain species.

In this study, I developed 18 chloroplast CDS gene-derived, species-specific SNP markers for each of seven *Panax* species including five representative *Panax* species and two basal *Panax* species in Asia and North America. Recently we developed *P. ginseng* cultivar-unique markers based on comprehensive comparative genomics analysis among chloroplast genome

sequences of 12 ginseng cultivars (Kim *et al.* 2015a). I excluded those intra-species polymorphic markers in this study because the aim of this study is to distinguish among different species. I also excluded the chloroplast genome targets which are transferred into mitochondrial genomes. All 18 dCAPS markers presented in this study are unique for one of the seven species and can be practically applied towards species authentication and breeding.

## REFERENCES

- Allen G, Flores-Vergara M, Krasynanski S, Kumar S, Thompson W (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols* 1:2320-2325
- Artyukova E, Kozyrenko M, Reunova G, Muzarok T, Zhuravlev YN (2000) RAPD analysis of genome variability of planted ginseng, *Panax ginseng*. *Molecular Biology* 34:297-302
- Chan T, But P, Cheng S, Kwok I, Lau F, Xu H (2000) Differentiation and authentication of *Panax ginseng*, *Panax quinquefolius*, and ginseng products by using HPLC/MS. *Analytical Chemistry* 72:1281-1287
- Chen X, Liao B, Song J, Pang X, Han J, Chen S (2013) A fast SNP identification and analysis of intraspecific variation in the medicinal *Panax* species based on DNA barcoding. *Gene* 530:39-43
- Cho IH (2012) Effects of *Panax ginseng* in neurodegenerative diseases. *Journal of Ginseng Research* 36:342
- Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, Yang TJ (2015) Complete Chloroplast Genome Sequence of Tartary Buckwheat (*Fagopyrum tataricum*) and Comparative Analysis with Common Buckwheat (*F. esculentum*). *PloS One* 10:e0125332
- Choi HI, Kim NH, Lee J, Choi BS, Do Kim K, Park JY, Lee SC, Yang TJ (2013) Evolutionary relationship of *Panax ginseng* and *P. quinquefolius* inferred from sequencing and comparative analysis of expressed sequence tags. *Genetic Resources and Crop Evolution* 60:1377-1387
- Choi HI, Kim NH, Kim JH, Choi BS, Ahn IO, Lee JS, Yang TJ (2011) Development of reproducible EST-derived SSR markers and assessment of genetic diversity in *Panax ginseng* cultivars and related species. *Journal of Ginseng Research* 35:399

- Court WE (2006) Ginseng: The Genus *Panax*. Taylor & Francis E-Library 15:13-21
- Cui Y, Qin S, Jiang P (2014) Chloroplast transformation of *Platymonas* (*Tetraselmis*) *subcordiformis* with the bar gene as selectable marker. PloS One 9:e98607
- Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genetics 15:138
- Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Research 16:805-814
- Gui S, Wu Z, Zhang H, Zheng Y, Zhu Z, Liang D, Ding Y (2016) The mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. Scientific Reports 6:30158
- Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genetics 6:e1000834
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ (2014) Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evolutionary Biology 14:151
- Jiang M, Liu J, Quan X, Quan L, Wu S (2016) Different chilling stresses stimulated the accumulation of different types of ginsenosides in *Panax ginseng* cells. Acta Physiol Plant 38:210
- Jung HJ, Choi H, Lim HW, Shin D, Kim H, Kwon B, Lee JE, Park EH, Lim CJ (2012) Enhancement of anti-inflammatory and antinociceptive actions of red ginseng extract by fermentation. Journal of Pharmacy and Pharmacology 64:756-762



- Jung J, Kim KH, Yang K, Bang KH, Yang TJ (2014) Practical application of DNA markers for high-throughput authentication of *Panax ginseng* and *Panax quinquefolius* from commercial ginseng products. *Journal of Ginseng Research* 38:123-129
- Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, Park HS, Yang TJ (2015a) Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS One* 10: e0117159
- Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS, Koh HJ, Waminal NE, Choi HI, Kim NH et al (2015b) Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific reports* 5:15655
- Kim K, Nguyen VB, Dong JZ, Wang Y, Park JY, Lee SC, Yang TJ (2017) Evolution of the Araliaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 *Panax*-related species. *Scientific reports* 7:4917
- Kim NH, Choi HI, Ahn IO, Yang TJ (2012) EST-SSR marker sets for practical authentication of all nine registered ginseng cultivars in Korea. *Journal of Ginseng Research* 36:298
- Kim NH, Choi HI, Kim KH, Jang W, Yang TJ (2014a) Evidence of genome duplication revealed by sequence analysis of multi-loci expressed sequence tag-simple sequence repeat bands in *Panax ginseng* Meyer. *Journal of Ginseng Research* 38:130-135
- Kim YJ, Jeon JN, Jang MG, Oh JY, Kwon WS, Jung SK, Yang DC (2014b) Ginsenoside profiles and related gene expression during foliation in *Panax ginseng* Meyer. *Journal of Ginseng Research* 38:66-72
- Kleine T, Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology* 60:115-138

- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Research* 19:1639-1645
- Lee YS, Park HS, Lee DK, Jayakodi M, Kim NH, Koo HJ, Lee SC, Kim YJ, Kwon SW, Yang TJ (2017) Integrated transcriptomic and metabolomic analysis of five *Panax ginseng* cultivars reveals the dynamics of ginsenoside biosynthesis. *Frontiers in Plant Science* 8:1048
- Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, depamphilis CW (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* 22:1948-1963
- Li R, Ma PF, Wen J, Yi TS (2013) Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. *PloS One* 8:e78568
- Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. *Biological Reviews* 90:157-166
- Ma KH, Dixit A, Kim YC, Lee DY, Kim TS, Cho EG, Park YJ (2007) Development and characterization of new microsatellite markers for ginseng (*Panax ginseng* CA Meyer). *Conservation Genetics* 8:1507-1509
- Ngan F, Shaw P, But P, Wang J (1999) Molecular authentication of *Panax* species. *Phytochemistry* 50:787-791
- Nguyen VB, Park HS, Lee SC, Lee J, Park JY, Yang TJ (2017) Authentication markers for five major *Panax* species developed via comparative analysis of complete chloroplast genome sequences. *Journal of Agricultural and Food Chemistry* 65:6298-6306

- Oh JY, Kim YJ, Jang MG, Joo SC, Kwon WS, Kim SY, Jung SK, Yang DC (2014) Investigation of ginsenosides in different tissues after elicitor treatment in *Panax ginseng*. *Journal of Ginseng Research* 38:270-277
- Park S, Ruhlman TA, Sabir JS, Mutwakil MH, Baeshen MN, Sabir MJ, Baeshen NA, Jansen RK (2014) Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. *BMC Genomics* 15:405
- Radad K, Gille G, Liu L, Rausch WD (2006) Use of ginseng in medicine with emphasis on neurodegenerative disorders. *Journal of Pharmacological Sciences* 100:175-186
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16:944-945
- Shi FX, Li MR, Li YL, Jiang P, Zhang C, Pan YZ, Liu B, Xiao HX, Li LF (2015) The impacts of polyploidy, geographic and ecological isolations on the diversification of *Panax* (Araliaceae). *BMC Plant Biology* 15:297
- Shi W, Wang Y, Li J, Zhang H, Ding L (2007) Investigation of ginsenosides in different parts and ages of *Panax ginseng*. *Food Chemistry* 102:664-668
- Smith DR, Crosby K, Lee RW (2011) Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. *Genome Biology and Evolution* 3:365-371
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. *American Journal of Botany* 96:336-348

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729
- Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* 5:123
- Wang X, Sakuma T, Asafu-Adjaye E, Shiu GK (1999) Determination of ginsenosides in plant extracts from *Panax ginseng* and *Panax quinquefolius* L. by LC/MS/MS. *Analytical Chemistry* 71:1579-1584
- Wong AS, Che CM, Leung KW (2015) Recent advances in ginseng as cancer therapeutics: a functional and mechanistic overview. *Natural Product Reports* 32:256-272
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* 106:13875-13879
- Xiao D, Yue H, Xiu Y, Sun X, Wang Y, Liu S (2015) Accumulation characteristics and correlation analysis of five ginsenosides with different cultivation ages from different regions. *Journal of Ginseng Research* 39:338-344
- Xie JT, Mehendale SR, Li X, Quigg R, Wang X, Wang CZ, Wu JA, Aung HH, A-Rue P, Bell GI, Yuan CS (2005) Anti-diabetic effect of ginsenoside Re in ob/ob mice. *Biochimica et Biophysica Acta* 1740:319-325
- Yang W, Qiao X, Li K, Fan J, Bo T, Guo DA, Ye M (2016) Identification and differentiation of *Panax ginseng*, *Panax quinquefolium*, and *Panax notoginseng* by monitoring multiple diagnostic chemical markers. *Acta Pharmaceutica Sinica B* 6:568-575

- Yuk J, McIntyre KL, Fischer C, Hicks J, Colson KL, Lui E, Brown D, Arnason JT (2013) Distinguishing Ontario ginseng landraces and ginseng species using NMR-based metabolomics. *Analytical and Bioanalytical Chemistry* 405:4499-4509
- Zheng SD, Wu HJ, Wu DL (2012) Roles and mechanisms of ginseng in protecting heart. *Chinese Journal of Integrative Medicine* 18:548-555
- Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou S (2011) DNA barcoding of *Panax* species. *Planta Medica* 77:182

## **CHAPTER III**

**Optimal condition and genotypes for transgenic hairy  
root induction and metabolite production in *P.*  
*ginseng***

## ABSTRACT

Ginseng is one of the most important medicinal plants containing a lot of pharmacologically valuable ginsenosides. However, the cultivation of ginseng takes about 4 - 6 years with extensive efforts to control quality from biotic and abiotic stresses. The *in vitro* tissue culture methods/systems have been developed to produce biomass for ginsenoside extraction, such as adventitious roots, cell suspension culture. However, these methods showed low biomass growth rate and normally go with plant hormone application, which increased the cost of products. In order to overcome such limitations, I have developed an efficient system for ginsenoside production through hairy root biomass mediated by *Agrobacterium rhizogenes* transformation. I successfully developed five transgenic hairy root lines, which can use for further biomass production. Among five transformed lines, Yunpoong cultivar and Ganghwa local landrace showed highest efficiency of transformation with 66.11% and 65.00%, respectively. PCR analysis of hairy roots confirmed the presence of *rolB* and *rolC* in the transformed lines, while these two genes were absent in normal adventitious roots. Transgenic hairy roots induced by *A. rhizogenes* grew 1.5 – 2.1 times faster in hormone-free SH medium than adventitious roots. Hairy roots produced the identical ginsenoside in comparison with those of adventitious roots, although total ginsenoside content were lower than those of adventitious roots, some ginsenosides, such as Rg<sub>1</sub>, Rf, Rh<sub>1</sub>, Rb<sub>1</sub>, Rb<sub>2</sub>, Rd, F<sub>2</sub>, and Rg<sub>3</sub> were biosynthesized in similar amounts or higher than those in adventitious roots. My result indicated that large scale production of hairy roots can be achieved by bioreactor culture and can be used for ginsenoside extraction.

**Keywords:** Plant transformation, hairy root, *P. ginseng*, *A. rhizogenes*, ginsenosides.

## INTRODUCTION

Ginseng, the king of herbal medicinal plants, have been used as traditional medicine for thousands of years in oriental countries owing to its unique pharmacological and economical value. To date, more than 200 ginsenosides, a triterpene saponins specifically biosynthesized in ginseng species and non-saponin constituents have been isolated in ginseng (Zhang *et al.* 2006). Among of *Panax* species, *P. ginseng* is known for containing relatively wider diversity of ginsenoside type (Kim *et al.* 2016). The global demand for herbal medicines is steadily increasing due to reinsured interest of consumers in natural products (Chen *et al.* 2016; Hishe *et al.* 2016; Verma and Singh 2008). However, the production and supply of herbal plants are still limited because of their growing conditions, overharvesting, and habitat destruction. In general, ginseng grows 4 - 6 years under the forest with a canopy cover up to 75 – 80% or in shade condition (Han *et al.* 2017; Pritts 2010). Further, in field cultivation, ginseng has to deal with a lot of difficulties including biotic and abiotic stresses (Hahn *et al.* 2001; Yu and Ohh 1994).

Various strategies and approaches have been developed using cell and organ plant cultures via biomass improvement for production of secondary metabolites in plants (Murthy *et al.* 2014). Similar techniques have also been developed to produce ginseng biomass in *in vitro* system for ginsenoside production, such as producing biomass of callus, suspension, and adventitious root (Ali *et al.* 2006; Furuya *et al.* 1984; Langhansova *et al.* 2005; Nguyen and Paek 2010; Woo *et al.* 2004; Yu *et al.* 2002). In addition, many propagation systems have been invented for biomass production, such as bioreactor, liquid shaking culture, and temporary immersion culture (Baque *et al.* 2012; Paek *et al.* 2005; Vaněk *et al.* 2005). However, all of these methods are still practically limited due to low growth rate of plant



tissues, and requirement of supplemental plant hormones for their growth. This led to increased cost of ginsenoside production.

Genetic engineering enables us to manipulate genes among species. The newly incorporated foreign genes in the host genome can bring new traits which can be observed in altered genotypes and/or phenotypes (Council 1987). Among different plant breeding techniques, a combination of transformation protocols with tissue culture seems to be promising and easy to produce novel homozygous breeding materials within a short span of time. Conventionally, the genetic engineering procedure of ginseng involves six major steps: 1) standardization of plant regeneration protocols, 2) desirable gene isolation from any species and vector construction, 3) transformation of gene into plants by *Agrobacterium* or by any other efficient method, 4) regeneration of putative transformed plants in selection media, 5) confirmation of transgene integration in ginseng genome and 6) characterization of transgenic plant phenotypes towards identification of new trait. *Agrobacterium rhizogenes* is a gram-negative soil bacterium responsible for hairy root induction using Ri T-DNA. Moreover, hairy root cultures are becoming attractive for expressing recombinant proteins (Ghiasi *et al.* 2012; Ono and Tian 2011). Hairy root induction by *A. rhizogenes* has many advantages including faster growth in hormone-free media, strong root branching, genetic and biosynthetic stability, plagiogeotropism, and biosynthetic capacity comparable to native plant roots (Pratap Chandran and Potty 2011; Tao and Li 2006). In the last decade, transformed root cultures from plants have attracted considerable attention because of their genetic and biochemical stability, rapid growth rate and their ability to synthesize secondary products at levels comparable to wild type roots (Giri and Narasu 2000).

In this study, I have demonstrated an efficient method for the development of hairy root culture as a biotechnological tool for secondary metabolite production in *P. ginseng*. Transgenic hairy roots have been generated via transformation mediated by *A. rhizogenes* harboring pBI121 binary vector. I also investigated transformation efficiency between *P. ginseng* cultivars, and produced transgenic hairy roots in liquid and bioreactor culture to select a valuable material for production of ginsenosides.

## MATERIALS AND METHODS

### Plant materials and tissue culture conditions

Callus of five *P. ginseng* accessions, 2 cultivars (Gumpoong and Yunpoong), 2 local landraces (Ganghwa and Jakyung) and 1 wild collections (Kangwon), were maintained and amplified in the basal MS medium (Murashige and Skoog 1962) supplement with 1.0 mg/L 2,4-D (2,4-Dichlorophenoxyacetic acid), 30 g/L sucrose, 7 g/L plant agar at  $23 \pm 1^{\circ}\text{C}$  in dark condition. The medium was adjusted to pH 5.8 before adding agar, and then sterilized at  $121^{\circ}\text{C}$  for 15 min by autoclave. After 30 days of subculture, these callus were used for transformation.

### *Agrobacterium* binary vector

The *A. rhizogenes* R1000 strain, provided by the bank of microbes, were used in this experiment. The binary vector pBI121 (Jefferson 1987), containing a CaMV 35S promoter-GUS gene fusion and the NPT II gene as a selectable marker, was transferred into *A. rhizogenes* R1000 by electroporation. Transformed *A. rhizogenes* were cultured in Luria-Bertani medium (1% tryptone, 0.5% yeast extract and 1% NaCl, pH 7.0) containing 50 mg/L kanamycin at  $28^{\circ}\text{C}$  with shaking (220 rpm) to mid-log phase ( $\text{OD}_{600} = 0.5$ ). The bacterial cells were collected by centrifugation at 3000 rpm for 10 min and resuspended in half MS liquid medium containing 30 g/L sucrose for inoculation.

### Plant transformation

Subcultured callus were excised by scalpel and directly used for co-cultivation with *A. rhizogenes*. The excised explants were dipped into the liquid inoculation medium containing suspension of *A. rhizogenes* for 10 min, then blotted dry using sterilized filter paper, and incubated in the dark at  $23 \pm 1^{\circ}\text{C}$  on half MS solid medium. After 2 days of co-cultivation, the explants were washed several times by sterilized water to remove *A. rhizogenes* on

the surface of explants, and blotted dry using sterilized filter paper. The co-cultivated explants were transferred to hormone-free half MS solid medium containing 250 mg/L cefotaxime to eliminate the residual bacteria, and incubated in the dark at  $23 \pm 1^\circ\text{C}$  for one week. After one week the explants were transferred to hormone-free half MS solid medium containing 30 g/L sucrose, 250 mg/L cefotaxime, and 50 mg/L kanamycin for selection of transgenic hairy roots. Putative transgenic hairy roots were observed emerging from the wound sites of explants after 4 weeks of incubation. Subsequently, the induced putative transgenic hairy roots were isolated and transferred to 30 ml of hormone-free Schenk and Hildebrandt (SH) liquid medium containing 30 g/L sucrose, 250 mg/L cefotaxime, and 50 mg/L kanamycin in 100 ml flask on a rotary shaker (100 rpm) at  $23 \pm 1^\circ\text{C}$  in the dark. Subculture were done after each 30 days.

#### **Histochemical GUS assay**

The hairy roots and adventitious roots were examined for histochemical GUS activity in staining solution (100 mM sodium phosphate, 1 mg/mL X-gluc, 0.1% Triton X-100, 10 mM EDTA, 2 mM potassium ferricyanide, and 2 mM potassium ferrocyanide). Hairy roots were soaked in the staining solution, and incubated overnight at  $37^\circ\text{C}$ . Stained roots were rinsed extensively in 70% ethanol to remove residual phenolic compounds, and observed using a microscope equipped with camera.

#### **Isolation of DNA and RNA, cDNA synthesis**

Total genomic DNA and RNA of hairy roots and adventitious roots were extracted using DNeasy and RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol, respectively. The cDNA were synthesized from 5  $\mu\text{g}$  of total RNA using SMARTer cDNA Synthesis Kit (Clontech Laboratories, Inc.) following the manufacturer's protocol.

#### **Confirmation of transformation by PCR analysis**

The polymerase chain reaction (PCR) was performed to confirm the presence of transgenes and introgressed *rol* genes. The oligonucleotide sequences used to amplify fragment of the GUS and *rol* genes are described in Table 3-1. The polymerase chain reaction (PCR) was performed in a 25  $\mu$ L reaction mixture containing 20 ng of DNA template, 5 pmol of each primer, 1.25 mM deoxynucleotide triphosphate (dNTP), 1.25 units of *Taq* DNA polymerase (Inclone, Korea), and 2.5  $\mu$ L of 10 $\times$  reaction buffer. The PCR reaction was performed in thermocyclers using the following cycling parameters: 94°C (5 min); 35 cycles of 94°C (30 s), 54–58°C (30 s); 72°C (30-60 s), then 72°C (7 min). PCR products were visualized on agarose gels (1.0 %) containing safe gel stain.

#### **Gene expression analysis by RT-PCR**

Expression of transgenes at the transcription level was analyzed by reverse transcriptase polymerase chain reaction. Synthesized cDNA was diluted 1/10, and used as template for PCR analysis. PCR condition and primers for GUS, *rol* and Actin7 genes were same as that used in the transformation confirmation and in Table 3-1.

#### **Liquid shaking culture and Bioreactor for biomass production**

Transgenic hairy roots, selected several times in hormone-free Schenk and Hildebrandt liquid medium containing 30 g/L sucrose, 250 mg/L cefotaxime, and 50 mg/L kanamycin, were transfer to hormone-free SH liquid medium containing 50 g/L sucrose without cefotaxime, and kanamycin for amplification. Subsequently, these hairy roots were used for growth rate comparison. 1 g fresh weigh of hairy roots were inoculated into 30 mL of Schenk and Hildebrandt (SH) liquid medium containing 50 g/L sucrose with or without 3 mg/L IBA in 100 ml flask with shaking (100 rpm) at 23  $\pm$  1°C in the dark. Data were collect after 30 days of culture.

Biomass production of transgenic hairy roots was performed in the air-lift balloon-type bioreactors. The bioreactor contained 1 L of the hormone-free SH liquid medium supplement with 5% sucrose. The airflow rate was adjusted at 0.1 vvm (100 mL/min) during the cultivation. The bioreactor was maintain in dark condition at  $23 \pm 1^\circ\text{C}$ . 12 g fresh weight of hairy roots was inoculated into a 3 L air-lift balloon-type bioreactor containing 1 L SH liquid medium for biomass production. Hairy root biomass were harvested after 30 days of culture.

### **Sample preparation for ginsenoside quantification**

Pulverized, freeze-dried adventitious and hairy roots (25 mg) were extracted with 1 mL of 70% methanol by sonication for 180 min at room temperature. The crude extract was centrifuged at 13,000 rpm for 5 min, then supernatant was diluted with water (1→3) and filtered using 0.2  $\mu\text{m}$  RC-membrane filter (Minisart RC15, Sartorius Stedim Biotech, Gottingen, Germany). The stock solutions of 13 ginsenoside reference standards were dissolved in MeOH. The stock solutions were mixed and diluted with MeOH to obtain a series of mixture solutions in desired concentrations. The solutions were filtered through a 0.2  $\mu\text{m}$  RC-membrane filter prior to quantitative analysis. All the solutions were stored at  $4^\circ\text{C}$  until analysis.

### **Ginsenoside quantification using LC–MS**

Quantification of ginsenosides were performed using an Agilent 6460 Triple Quadrupole LC/MS system (Agilent Technologies, Palo Alto, CA, USA) coupled with an Agilent 1290 Infinity II LC (Agilent Technologies). Chromatographic separation was obtained with an ACQUITY UPLC BEH C18 column (100 mm  $\times$  2.1 mm, 1.7  $\mu\text{m}$ ; Waters Co., Milford, MA, USA) operated at  $40^\circ\text{C}$ . The mobile phase consisting of  $\text{H}_2\text{O}$  (A) and acetonitrile (B), both of which were added with 0.1% formic acid, was delivered at a flowrate of 0.3 mL/min with a gradient program: 0–9 min, 20% (B); 9–14

min, 20–30% (B); 14–17 min, 30% (B); 17–21 min, 30–32% (B); 21–26 min, 32–42.5% (B); 26–29 min, 42.5–90%; 29–31 min, 90% (B); 31–32 min, 90–20%; 32–35 min, 20% (B). The sample injection volume was set to 2.0  $\mu$ L. The triple quadrupole MS system was equipped with an electrospray ionization (ESI) interface (ESI) and operated in the negative ion mode. Parameters for MS were set as followings: capillary voltage 3.5 kV; Vcharge 2.0 kV; drying and sheath gas temperature 320 °C; drying gas flow 10 L/min; sheath gas flow 11 L/min; nebulizer pressure 50 psi.

### **Statistical Analysis**

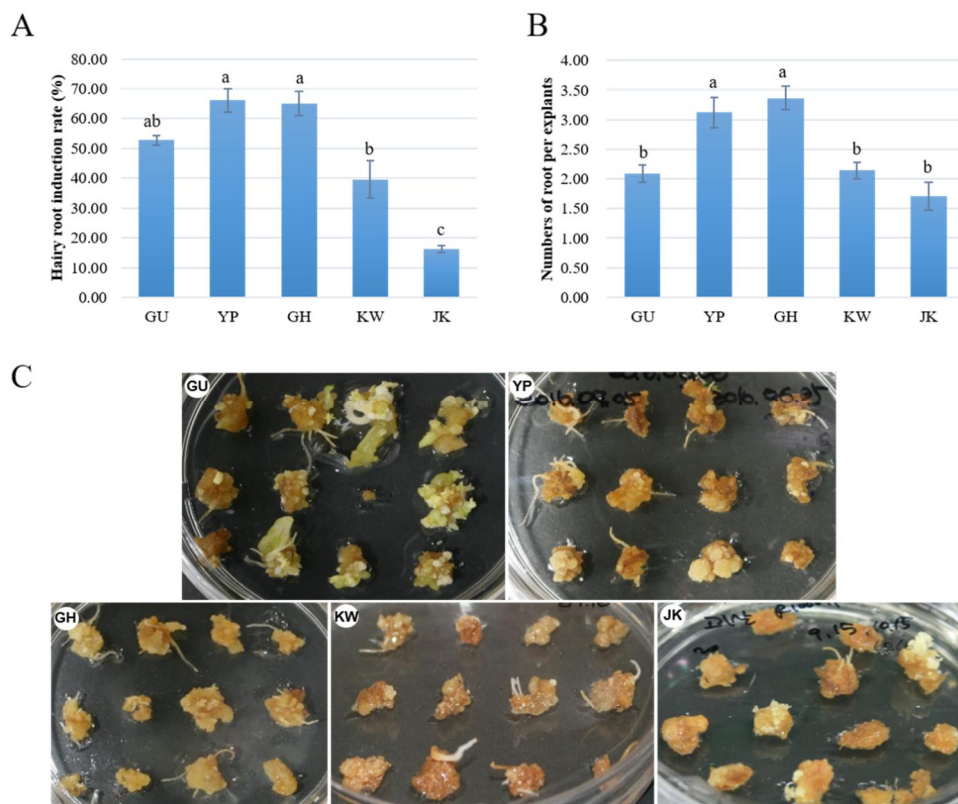
All statistical analysis was conducted using R language (version 3.4.0). Data were analyzed by one-way analysis of variance (ANOVA) and significant differences among values were performed by the Tukey's multiple comparison test ( $P \leq 0.05$ ).

## RESULTS

### **Efficiency of *A. rhizogenes*-mediated transformation among five *P. ginseng* accessions**

Earlier studies showed that several factors likewise genotype, explants, physical and chemical factors, strain of *A. rhizogenes* can affect genetic transformation induced by *A. rhizogenes* (Kumar *et al.* 1991; Sharafi *et al.* 2014). I found that the transformation efficiency was relative high in all 5 samples from 16.33% - 66.11%, and that transformation efficiency of Yunpoong cultivar and Ganghwa local landrace was significantly higher than that of Kangwon, Jakyung, but not significantly different with that of Gumpoong (Fig. 3-1A, C). Number of hairy roots per explants was also different among samples from 1.71 – 3.36 roots, and these root number per explants were significantly higher in Yunpoong and Ganghwa than those of Gumpoong, Kangwon, and Jakyung (Fig. 3-1B, C). My result indicated that callus was suitable explant for *A. rhizogenes* transformation, and among 5 *P. ginseng* lines, Yunpoong and Ganghwa were the best candidate for *A. rhizogenes* transformation.





**Figure 3-1. Efficiency of *Agrobacterium rhizogenes* transformation of five *P. ginseng* accessions.** (A) Hairy root induction rate, (B) number of root per explant, (C) hairy root induction from transformed callus of five *P. ginseng* accessions on solid selection media. GU, *P. ginseng* cv. Gumpoong; YP, *P. ginseng* cv. Yunpoong; GH, *P. ginseng* local landrace Ganghwa; KW, *P. ginseng* wild collection Kangwon; JK, *P. ginseng* local landrace Jakyung.

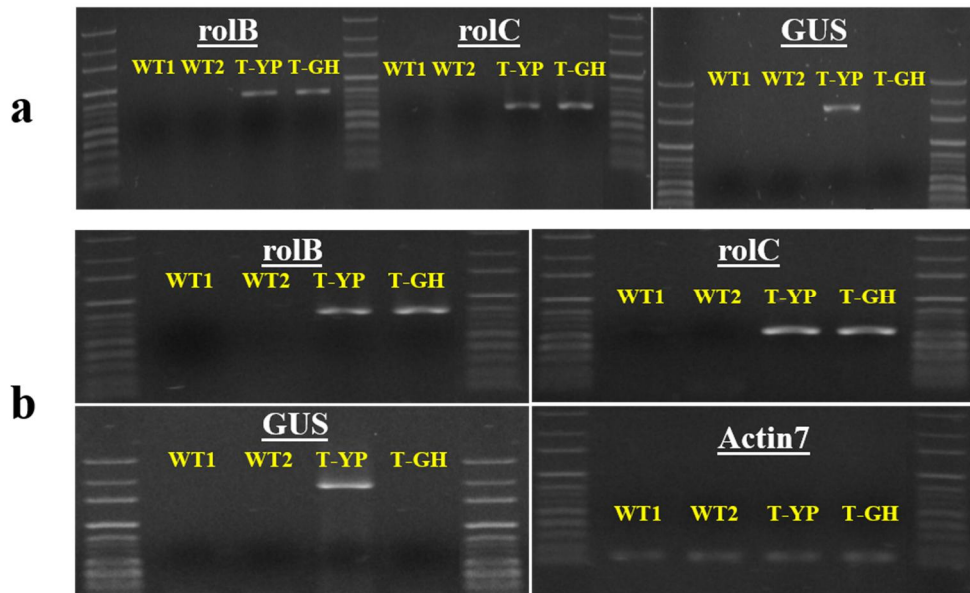
### **Molecular analysis of transgenic hairy roots**

I selected putative hairy roots from 2 representative lines (Yunpoong and Ganghwa) for molecular verification and next experiments. PCR-based analysis of *rolB* and *rolC* gene was performed to confirm the genetic transformation of selected hairy root lines (Table 3-1). The *rol* genes of Ri-plasmid are responsible for hairy root induction by *A. rhizogenes*, and were used to check the presence of these fragments into the genomic DNA of transgenic hairy root. The presence of *rolB* and *rolC* gene in PCR amplification were observed in the established hairy root lines indicating that the pRi T-DNA fragments of *A. rhizogenes* was successfully integrated into the plant genome for hairy root inducing (Fig. 3-2a). No amplification was observed in using DNA extracted from adventitious roots, the non-transform roots (Fig. 3-2a). Additionally, I checked the GUS gene to confirm the complete-induced hairy roots. Interestingly, I found that the presence of GUS gene in hairy roots of Yunpoong, but not in Ganghwa hairy roots (Fig. 3-2a). To confirm hairy roots were induced by integrated of *rol* genes from Ri-plasmid, I performed RT-PCR to check the expression of *rol B* and *rol C* genes at transcription level. I also confirmed the expression of GUS genes from binary vector in all hairy roots. RT-PCR results showed that *rol* genes were high expression in hairy root (Fig. 3-2b). GUS gene expression was also observed only in the hairy roots of Yunpoong (Fig. 3-2b). The GUS gene was not found in the hairy roots of Ganghwa, and this may cause by the loss or elimination of transgenes into plant cell of Ganghwa.

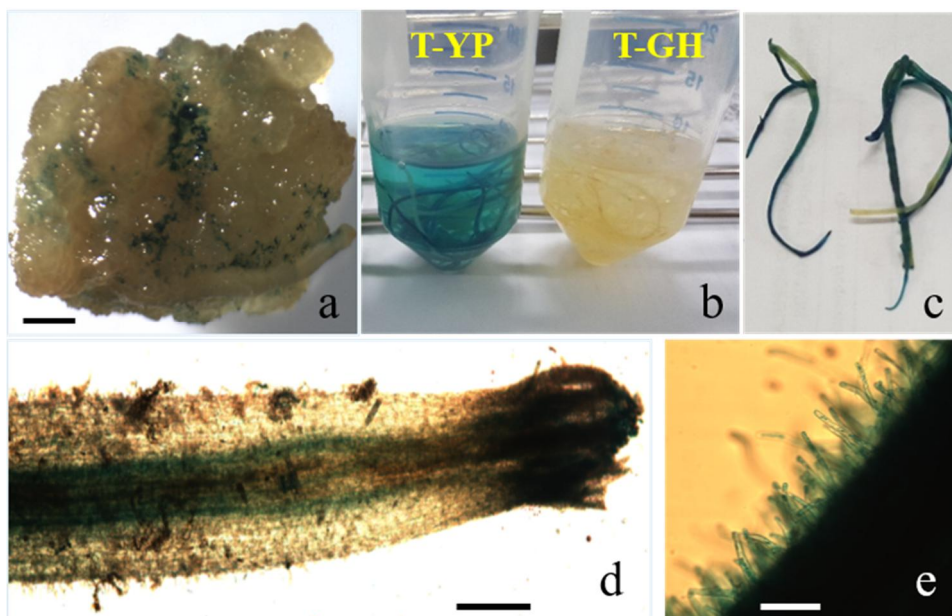
Histochemical staining for GUS activity was performed to determine whether *A. rhizogenes* strain R1000 produced complete root transformation or not. The strong GUS staining was observed in all transgenic hairy roots of Yunpoong inducing from *A. rhizogenes* strain R1000 (pBI121), but not in Ganghwa hairy roots (Fig. 3-3).

**Table 3-1.** Primer information for PCR and RT-PCR analysis.

<b>Target gene</b>	<b>Direction</b>	<b>Sequence (5'-&gt;3')</b>	<b>Melting temperature (°C)</b>
<b>RolB</b>	<b>F</b>	<b>GCTCTTGCAGTGCTAGATTT</b>	<b>53.0</b>
	<b>R</b>	<b>GAAGGTGCAAGCTACCTCTC</b>	<b>55.3</b>
<b>RolC</b>	<b>F</b>	<b>ATGGCTGAAGACGACCTGTGT</b>	<b>58.7</b>
	<b>R</b>	<b>TTAGCCGATTGCAAACCTTGCA</b>	<b>55.8</b>
<b>GUS</b>	<b>F</b>	<b>ATGTTACGTCCTGTAGAAACCCC</b>	<b>56.5</b>
	<b>R</b>	<b>TCATTGTTTGCCTCCCTGCTGC</b>	<b>60.6</b>
<b>Actin7</b>	<b>F</b>	<b>CTTGAGACCTCAAAGACTAGAC</b>	<b>52.2</b>
	<b>R</b>	<b>TCTCGTGAATTCCTGCAGCT</b>	<b>56.5</b>



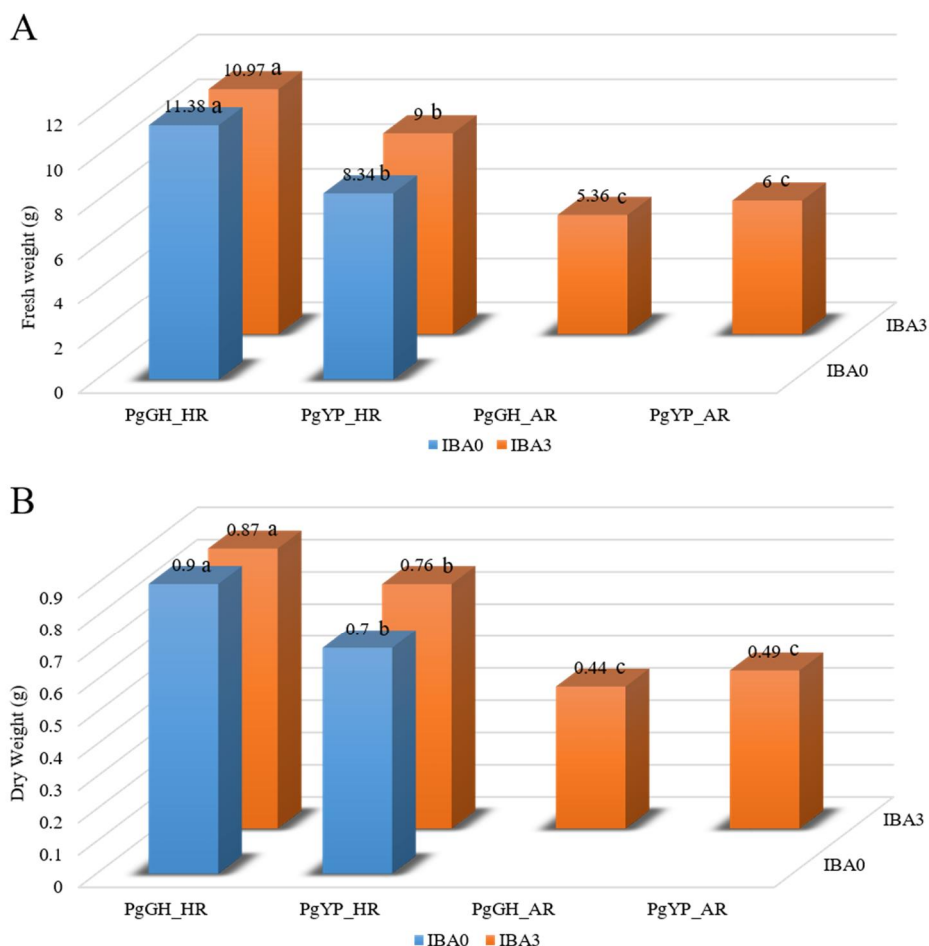
**Figure 3-2. Analysis in transgenic hairy roots of ginseng.** (a) PCR analysis for *rolB*, *rolC*, and *GUS* genes in transgenic hairy root. (b) Expression of *rolB*, *rolC*, *GUS* and *Actin7* at transcription level. T-YP, transgenic hairy roots of Yunpoong; T-GH, transgenic hairy roots of Ganghwa. WT1, adventitious roots of Yunpoong; WT2, adventitious roots of Ganghwa.



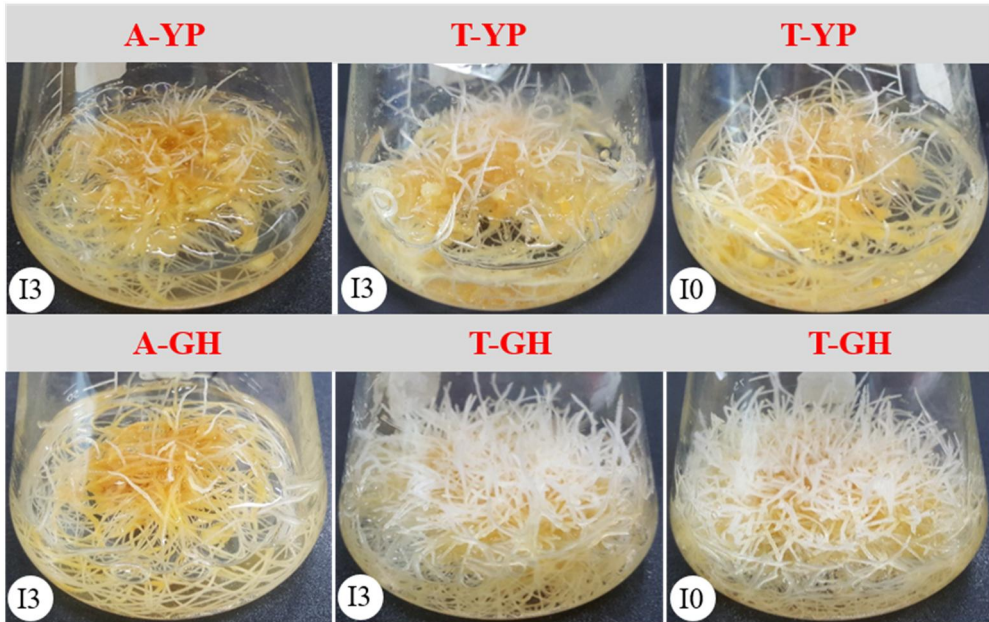
**Figure 3-3. GUS histochemical analysis of transgenic hairy roots of ginseng transformed by *A. rhizogenes* R100 (pBI121).** (a) Transient GUS spots on callus (bar = 2 mm), (b-d) transgenic hairy root after two months on selection medium (bar = 2 mm), (e) lateral root hair of hairy root (bar = 0.2 mm). T-YP, transgenic hairy roots of Yunpoong; T-GH, transgenic hairy roots of Ganghwa.

### **Comparison of hairy root growth rate in liquid shaking culture**

To investigate the growth rate of hairy roots, I cultured hairy roots in both medium with and without 3 mg/L IBA. First, the growth rate of hairy roots in both medium with and without IBA were faster and significantly higher than those of adventitious roots in medium supplement with 3 mg/L IBA, and these were clearly showed by fresh and dry weight (Fig. 3-4A, B). Regarding to morphology, transgenic hairy roots with a lot of lateral root hairs were also different with adventitious roots (Fig. 3-5). Interestingly, I found that the growth rate of transgenic hairy roots in medium with and without 3 mg/L IBA were not different both in fresh weigh or dry weigh status (Fig. 3-4A, B). This result indicated that hairy root growth rate were not affected by auxin (IBA) application, and hairy root can grow in medium without auxin (IBA). In comparison of growth rate between 2 transgenic lines, I found that the hairy root growth rate of Ganghwa were higher than that of Yunpoong both in fresh and dry weight (Fig. 3-4A, B).



**Figure 3-4. Growth rate of transgenic hairy roots in liquid medium. (A)** Fresh weight of hairy roots and adventitious roots in medium with or without 3 mg/L IBA. **(B)** Dry weight of hairy roots and adventitious roots in medium with or without 3 mg/L IBA. PgYP\_HR, PgGH\_HR: hairy roots of Yunpoong and Ganghwa; PgYP\_AR and PgGH\_AR: adventitious roots of Yunpoong and Ganghwa, respectively.



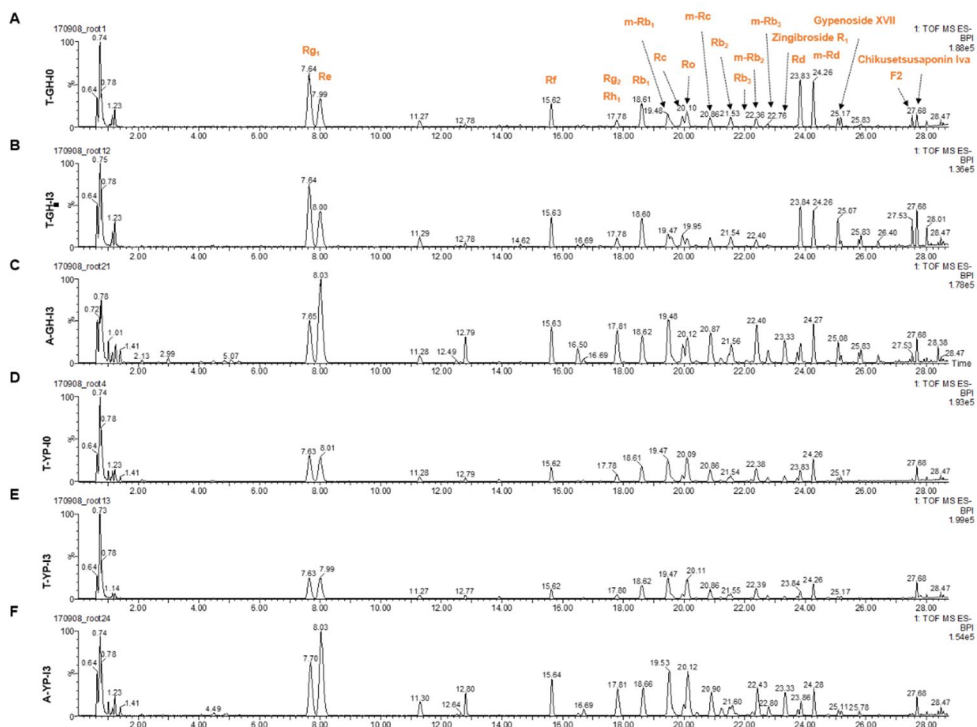
**Figure 3-5. Morphology of adventitious roots and transgenic hairy roots in liquid medium.** I3, SH medium supplement with 3 mg/L IBA; I0, IBA-free SH medium. A-YP, adventitious roots of Yunpoong; A-GH, adventitious roots of Ganghwa. T-YP, transgenic hairy roots of Yunpoong; T-GH, transgenic hairy roots of Ganghwa.



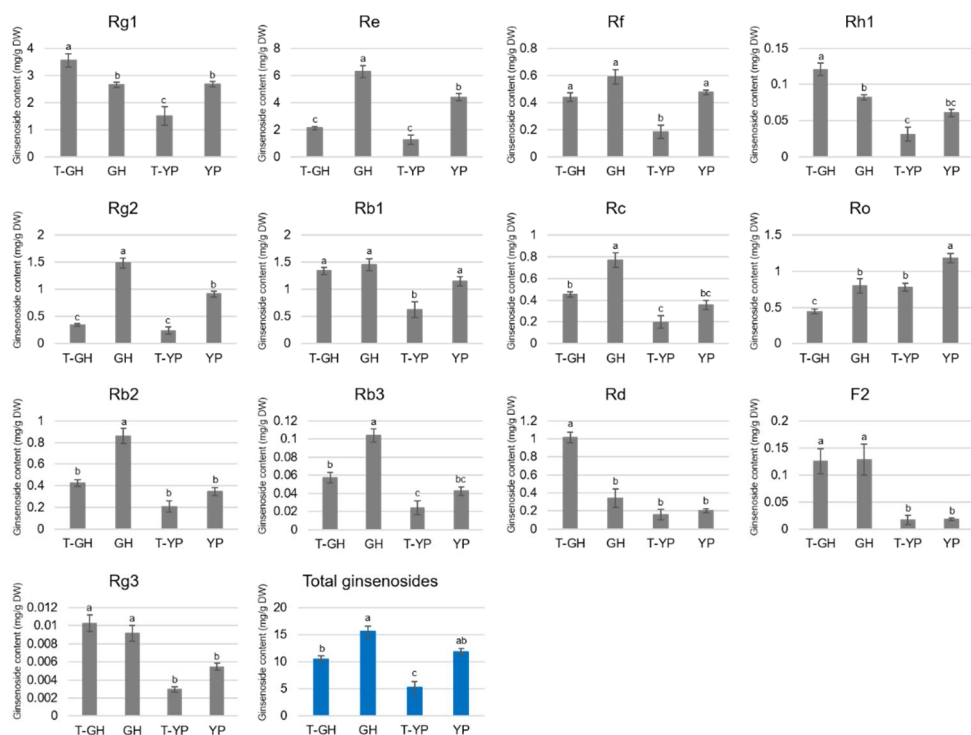
### **Ginsenoside profiling in hairy roots compared to adventitious roots**

I performed secondary metabolite profiling for two transgenic hairy root lines (Ganghwa and Yunpoong) in comparison with their adventitious roots using **UPLC-QTOP/MS**. The results indicated that secondary metabolite pattern were similar between transgenic hairy roots and adventitious roots (Fig. 3-6). All common ginsenosides such as Rb<sub>1</sub>, Rb<sub>2</sub>, Rc, Rd, Re, Rg<sub>1</sub>, Rg<sub>2</sub>, Rf were identified in transgenic hairy roots, and showed the similarity with that of adventitious roots (Fig. 3-6)

I quantified contents of 13 ginsenosides, Rb<sub>1</sub>, Rb<sub>2</sub>, Rb<sub>3</sub>, Rc, Rd, Re, Rf, Rh<sub>1</sub>, Rg<sub>1</sub>, Rg<sub>2</sub>, Rg<sub>3</sub>, Ro, and F<sub>2</sub>, in two transgenic hairy root lines (Ganghwa and Yunpoong) and their adventitious roots using LC-ESI-MS/MS in multiple reaction monitoring (MRM) acquisition mode. The relative abundance ratio between each ginsenoside was similar in both of adventitious roots and hairy roots (Fig. 3-7). However, for both of Yunpoong and Ganghwa, hairy roots tended to show lower contents of ginsenosides than adventitious roots. This was especially remarkable for hairy roots of Yunpoong cultivar, in which almost every ginsenosides showed lower concentration than those in adventitious roots except the equal contents of Rb<sub>2</sub>, Rd, F<sub>2</sub> and Rg<sub>3</sub> ginsenoside (Fig. 3-7). In the case of Ganghwa local landrace, hairy roots were rich in Rg<sub>1</sub>, Rh<sub>1</sub>, and Rd ginsenosides, while adventitious roots were rich in Re, Rg<sub>2</sub>, Rc, Rb<sub>2</sub>, Rb<sub>3</sub>, and Ro ginsenosides (Fig. 3-7). The other ginsenosides such as Rf, Rb<sub>1</sub>, F<sub>2</sub> and Rg<sub>3</sub> were similar between hairy roots and adventitious roots (Fig. 3-7). In comparison between Ganghwa local landrace and Yunpoong cultivars, Ganghwa showed higher amounts of total ginsenosides than those of Yunpoong both in hairy roots and adventitious roots, respectively (Fig. 3-7).



**Figure 3-6. UPLC-QTOP/MS chromatogram of adventitious roots and transgenic hairy roots. (A-C) transgenic and adventitious roots of Ganghwa, (D-F) transgenic and adventitious roots of Yunpoong.**



**Figure 3-7. Quantitative analysis of 13 major ginsenosides in adventitious roots and transgenic hairy roots.** T-YP, transgenic hairy roots of Yunpoong; T-GH, transgenic hairy roots of Ganghwa; GH, adventitious roots of Ganghwa; YP, adventitious roots of Yunpoong.

### **Biomass production of hairy roots in bioreactor**

I applied bioreactor for hairy root production in the hormone-free SH medium supplement with 5% sucrose. The results showed that after 30 days of inoculation in bioreactor, total of 141.33 g hairy roots were harvested from initial amount of 12 g fresh weight, the biomass was increase 11.75 and 11.77 times in fresh weight and dry weight test, respectively (Fig. 3-8a, b). Analysis of ginsenosides in hairy roots produced in bioreactor system showed that hairy root contained all kind of ginsenoside as nature roots and adventitious in liquid shaking culture (Fig. 3-8c). This results indicated that bioreactor system is the best system, and hairy root is the best candidate material for biomass and ginsenoside production in industrial scale.



## DISCUSSION

Production of secondary metabolites from plants has been considered as promising alternative method due to many advantages such as: low cost, safety to human health, simple and easy for large-scale industrial application. Many economically important secondary metabolites for flavours, colours and pharmaceutical usage have been produced from plants (Hussain *et al.* 2012). In order to obtain high yields suitable for commercial exploitation, beside improving method, techniques or optimizing culture condition, transgenic plants have revolutionized the role of plant tissue culture in secondary metabolite production (Hussain *et al.* 2012). In this study, I developed a new transgenic hairy root resource which can be used for biomass production via bioreactor system, and success of this study provides a new material for large scale production of ginsenoside in *in vitro*.

There are lot of methods to generated transgenic plant but the transformation mediated by *A. rhizogenes* is considered as natural genetic engineer (Brijwal and Tamta 2015). *A. rhizogenes* mediated transformation is also the most common and effective method, by inducing proliferative root at the infection site caused by T-DNA (*rol* genes) insertion into plant genome. *A. rhizogenes* strains R1000 was reported as the best strain for hairy root induction in buckwheat (Thwe *et al.* 2016), watercress (Park *et al.* 2011), and radish (Bae *et al.* 2012). My results also showed that the infection efficiency of *A. rhizogenes* strains R1000 in ginseng were up to 66.11% (Fig. 3-1A). This indicated that *A. rhizogenes* strains R1000 is more reliable and convenient for transformation in ginseng. Plant tissue explants have been reported to effect transformation efficiency, and suitable explants selection were played an important role in successful transformation (Ismael and Antar 2014). Many plant tissue explants have been used for transformation by

*Agrobacterium* such as, hypocotyl, leaves, cotyledons (Aarrouf *et al.* 2012; Bae *et al.* 2012), adventitious root (Han and Choi 2009), and callus (Gorpenchenko *et al.* 2006; Ismael and Antar 2014). Callus tissue have been reported as an ideal material for transformation in plants, because of many advantages such as high infection of *Agrobacteria*, easy selection and regeneration (Al Abdallat *et al.* 2011; Ismael and Antar 2014; Komari 1989). Herein, I used callus of the five *P. ginseng* lines as explants for transformation and compared the efficiency of transformation between them to select the good material for future studies and biomass production. I obtained the high transformation efficiency for almost five lines (Fig. 3-1A), and found that transformed callus were not only induced hairy root but also induced somatic embryogenesis, and shoot organogenesis (Fig. 3-1C), which is consistent with the previous report from *Agrobacterium* (Gorpenchenko *et al.* 2006).

Transformed cells/tissue selection is the key step for transformation. This is facilitated by the present of selectable marker genes available in fusion vectors, which is usually an antibiotic or herbicide resistance gene. The common selection marker is an antibiotic resistance gene such as kanamycin. Reporter genes have been also offered as an alternative technique for selection, in which the most commonly used reporter genes are *nptII*, *cat*, *gus*, *lux* and *gfp*. In this study, I used the binary vector pBI121, containing GUS reporter gene and the NPT II gene as a selectable marker. I found that all representative transformed lines showed resistance to kanamycin, and hairy root can growth well in the medium supplement with 50 mg/l kanamycin. Additionally, PCR and RT-PCR results clearly showed the represent of *rol* gene in all transformed root. Moreover, based on root morphology and growing condition I can confirm whether these are really hairy root. Interestingly, for GUS assay I found that among the two

representative lines, just only Yunpoong showed GUS signal (Fig. 3-2, 3). The GUS staining results were in consistent with PCR and RT-PCR results. The absence of GUS gene in Ganghwa hairy roots may be caused by loss or elimination of transgenes, and this phenomenon was also reported in some transgenic plant (Hodal *et al.* 1992; Risseuw *et al.* 1997; Romano *et al.* 2005).

Production of plant biomass *in vitro* normally goes with the application of plant hormones, however, this application leads to increase the cost of products. In this study, I found that both hairy root lines were able to grow faster in hormone-free medium, and the growth rate were from 1.5 - 2.1 times higher than adventitious root (Fig. 3-4, 5). Comparison of ginsenosides showed that hairy root produced the same saponins, ginsenoside varieties as those of the adventitious root, and cultivated roots (Fig. 3-6). The difference of ginsenoside content between *P. ginseng* cultivars had been reported in previous study (Lee *et al.* 2017). Here, I found that Ganghwa local landrace showed higher ginsenoside content than that of Yunpoong cultivar both in hairy roots and adventitious roots (Fig. 3-7). Transgenic hairy roots tended to contain lower amounts of ginsenosides than those of adventitious roots (Fig. 3-7). However, some major ginsenosides showed higher or similar amounts in hairy roots in comparison with those in adventitious roots including Rg<sub>1</sub>, Rh<sub>1</sub> and Rd contents were higher, and Rf, Rb<sub>1</sub>, F2 and Rg<sub>3</sub> were similar in Ganghwa local landrace line (Fig. 3-7). Thus, hairy roots can be an efficient platform for production of these kinds of ginsenosides. Bioreactor system has been successfully applied for large scale production of many horticultural and medicinal plants (Paek *et al.* 2005). Application of bioreactor technology is the key step toward commercial production of bioactive molecules in plant and animal cell (Baque *et al.* 2012; Warnock and Al-Rubeai 2006). In this study, I applied bioreactor system for biomass



production of hairy roots, and obtained high amount and quality of hairy roots just in one month in free-hormone medium (Fig. 3-8). My study results indicated that hairy roots are the ideal material, and bioreactor is the effective system for *in vitro* biomass and secondary metabolite production.

## REFERENCES

- Aarrouf J, Castro-Quezada P, Mallard S, Caromel B, Lizzi Y, Lefebvre V (2012) *Agrobacterium rhizogenes*-dependent production of transformed roots from foliar explants of pepper (*Capsicum annuum*): a new and efficient tool for functional analysis of genes. *Plant Cell Reports* 31:391-401
- Al Abdallat A, Sawwan J, Al Zoubi B (2011) *Agrobacterium tumefaciens*-mediated transformation of callus cells of *Crataegus aronia*. *Plant Cell, Tissue and Organ Culture (PCTOC)* 104:31-39
- Ali MB, Yu KW, Hahn EJ, Paek KY (2006) Methyl jasmonate and salicylic acid elicitation induces ginsenosides accumulation, enzymatic and non-enzymatic antioxidant in suspension culture *Panax ginseng* roots in bioreactors. *Plant Cell Reports* 25:613-620
- Bae H, Kim YB, Park NI, Kim HH, Kim YS, Lee MY, Park SU (2012) *Agrobacterium rhizogenes*-mediated genetic transformation of radish (*Raphanus sativus* L. cv. Valentine) for accumulation of anthocyanin. *Plant Omics* 5:381
- Baque MA, Moh SH, Lee EJ, Zhong JJ, Paek KY (2012) Production of biomass and useful compounds from adventitious roots of high-value added medicinal plants using bioreactor. *Biotechnology Advances* 30:1255-1267
- Brijwal L, Tamta S (2015) *Agrobacterium rhizogenes* mediated hairy root induction in endangered *Berberis aristata* DC. *SpringerPlus* 4:443
- Chen SL, Yu H, Luo HM, Wu Q, Li CF, Steinmetz A (2016) Conservation and sustainable use of medicinal plants: problems, progress, and prospects. *Chinese Medicine* 11:37
- Council NR (1987) *Agricultural biotechnology: strategies for national competitiveness*. National Academies Press

- Furuya T, Yoshikawa T, Orihara Y, Oda H (1984) Studies of the culture conditions for *Panax ginseng* cells in jar fermentors. *Journal of Natural Products* 47:70-75
- Ghiasi SM, Salmanian AH, Sharafi A, Kazemi R, Jafari M, Chinikar S, Zakeri S (2012) Molecular farming, an effective system for the production of immunogenic crimean-congo hemorrhagic fever virus glycoprotein. *Progress in Biological Sciences* 2:12-29
- Giri A, Narasu ML (2000) Transgenic hairy roots: recent trends and applications. *Biotechnology Advances* 18:1-22
- Gorpenchenko T, Kiselev K, Bulgakov V, Tchernoded G, Bragina E, Khodakovskaya M, Koren O, Batygina T, Zhuravlev YN (2006) The *Agrobacterium rhizogenes* *rolC*-gene-induced somatic embryogenesis and shoot organogenesis in *Panax ginseng* transformed calluses. *Planta* 223:457-467
- Hahn E, Paek K, Yu K (2001) Ginsenoside production by hairy root cultures of *Panax ginseng* CA Meyer in bioreactors. *International Conference on Medicinal and Aromatic Plants (Part II)* 597: 237-243
- Han H, Jang KM, Chung JS (2017) Selecting suitable sites for mountain ginseng (*Panax ginseng*) cultivation by using geographically weighted logistic regression. *Journal of Mountain Science* 14:492-500
- Han JY, Choi YE (2009) Rapid induction of *Agrobacterium tumefaciens*-mediated transgenic roots directly from adventitious roots in *Panax ginseng*. *Plant Cell, Tissue and Organ Culture (PCTOC)* 96:143
- Hishe M, Asfaw Z, Giday M (2016) Review on value chain analysis of medicinal plants and the associated challenges. *Journal of Medicinal Plants Studies* 4:45-55
- Hodal L, Bocharadt A, Nielsen JE, Mattsson O, Okkels FT (1992) Detection, expression and specific elimination of endogenous  $\beta$ -glucuronidase

- activity in transgenic and non-transgenic plants. *Plant Science* 87:115-122
- Hussain MS, Fareed S, Saba Ansari M, Rahman A, Ahmad IZ, Saeed M (2012) Current approaches toward production of secondary plant metabolites. *Journal of Pharmacy & Bioallied Sciences* 4:10
- Khatab AI, El-Banna AN (2014) Establishment of high-efficiency *Agrobacterium*-mediated transformation conditions of soybean callus. *Indian Journal of Biotechnology* 13:459-463
- Jefferson RA (1987) Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Molecular Biology Reporter* 5:387-405
- Kim YJ, Silva J, Zhang D, Shi J, Joo SC, Jang MG, Kwon WS, Yang DC (2016) Development of interspecies hybrids to increase ginseng biomass and ginsenoside yield. *Plant Cell Reports* 35:779-790
- Komari T (1989) Transformation of callus cultures of nine plant species mediated by *Agrobacterium*. *Plant Science* 60:223-229
- Kumar V, Jones B, Davey M (1991) Transformation by *Agrobacterium rhizogenes* and regeneration of transgenic shoots of the wild soybean *Glycine argyrea*. *Plant Cell Reports* 10:135-138
- Langhansova L, Marsik P, Vanek T (2005) Production of saponins from *Panax ginseng* suspension and adventitious root cultures. *Biologia Plantarum* 49:463-465
- Lee YS, Park HS, Lee DK, Jayakodi M, Kim NH, Koo HJ, Lee SC, Kim YJ, Kwon SW, Yang TJ (2017) Integrated transcriptomic and metabolomic analysis of five *Panax ginseng* cultivars reveals the dynamics of ginsenoside biosynthesis. *Frontiers in Plant Science* 8:1048
- Murthy HN, Lee EJ, Paek KY (2014) Production of secondary metabolites from cell and organ cultures: strategies and approaches for biomass

- improvement and metabolite accumulation. *Plant Cell, Tissue and Organ Culture (PCTOC)* 118:1-16
- Nguyen TT, Paek KY (2010) Cell suspension culture *Panax ginseng* CA Meyer: Role of plant growth regulators and medium composition on biomass and ginsenoside production. *Natural Sciences and Technology* 26:191-196
- Ono NN, Tian L (2011) The multiplicity of hairy root cultures: prolific possibilities. *Plant Science* 180:439-446
- Paek K, Chakrabarty D, Hahn E (2005) Application of bioreactor systems for large scale production of horticultural and medicinal plants. *Plant Cell, Tissue Organ Culture* 81:287-300
- Park NI, Kim JK, Park WT, Cho JW, Lim YP, Park SU (2011) An efficient protocol for genetic transformation of watercress (*Nasturtium officinale*) using *Agrobacterium rhizogenes*. *Molecular Biology Reports* 38:4947-4953
- Pratap Chandran R, Potty V (2011) Initiation of hairy roots from *Canavalia* sp. using *Agrobacterium rhizogenes* 15834 for the co-cultivation of Arbuscular mycorrhizal fungi, *Glomus microcarpum*. *Journal of Agricultural Technology* 7:235-245
- Pritts KD (2010) Ginseng: how to find, grow, and use America's forest gold. Stackpole Books. Mechanicsburg, PA. 150 pp
- Risseeuw E, Dijk ME, Hooykaas PJ (1997) Gene targeting and instability of *Agrobacterium* T-DNA loci in the plant genome. *The Plant Journal* 11:717-728
- Romano E, Soares A, Proite K, Neiva S, Grossi M, Faria JC, Rech EL, Aragão FJ (2005) Transgene elimination in genetically modified dry bean and soybean lines. *Genetic and Molecular Research* 4:177-184

- Sharafi A, Sohi HH, Azadi P, Sharafi AA (2014) Hairy root induction and plant regeneration of medicinal plant *Dracocephalum kotschyi*. *Physiology and Molecular Biology of Plants* 20:257-262
- Tao J, Li L (2006) Genetic transformation of *Torenia fournieri* L. mediated by *Agrobacterium rhizogenes*. *South African Journal of Botany* 72:211-216
- Thwe A, Arasu MV, Li X, Park CH, Kim SJ, Al-Dhabi NA, Park SU (2016) Effect of different *Agrobacterium rhizogenes* strains on hairy root induction and phenylpropanoid biosynthesis in tartary buckwheat (*Fagopyrum tataricum* Gaertn). *Frontiers in Microbiology* 7:318
- Vaněk T, Langhansová L, Maršík P (2005) Cultivation of root cultures of *Panax ginseng* in different bioreactors and in temporary immersion-Comparison of growth and saponin production. In: Hvoslef-Eide A, Preil W (Eds) *Liquid culture systems for in vitro plant propagation*, Springer, The Netherlands, pp 539–547
- Verma S, Singh S (2008) Current and future status of herbal medicines. *Veterinary World* 1:347-350
- Warnock JN, Al-Rubeai M (2006) Bioreactor systems for the production of biopharmaceuticals from animal cells. *Biotechnology and Applied Biochemistry* 45:1-12
- Woo SS, Song JS, Lee JY, In DS, Chung HJ, Liu JR, Choi DW (2004) Selection of high ginsenoside producing ginseng hairy root lines using targeted metabolic analysis. *Phytochemistry* 65:2751-2761
- Yu KW, Gao W, Hahn EJ, Paek KY (2002) Jasmonic acid improves ginsenoside accumulation in adventitious root culture of *Panax ginseng* CA Meyer. *Biochemical Engineering Journal* 11:211-215
- Yu Y, Ohh S (1994) Problems and present status of research on ginseng diseases in Korea. *Proc Int Ginseng Conf Vancouver*, pp 120-130

Zhang J, Chui D, Chen C, Liu G (2006) The chemistry, metabolism and biological activities of ginseng. Beijing: Chemical Industry Press

## ABSTRACT IN KOREAN

인삼속은 두릅나무과에 속하며 다양한 약리작용을 보이는 진세노사이드가 풍부하여 허브의 왕으로 불리고 있다. 현재까지 14 종의 인삼속 식물들이 보고되어 있으며 이 중 인삼, 미국삼, 전칠삼, 죽절삼, 베트남삼 5 종이 대표적이며 한국, 미국, 캐나다, 중국, 일본, 베트남에서 널리 사용되고 있다. 그러나 이들의 유전적 다양성과 진화적 유연관계, 그리고 기원에 관해서는 밝혀진 바가 부족하다. 각각의 인삼속 식물은 효용성과 가격적인 측면에서 차이가 있지만 그럼에도 불구하고 약리적, 경제적인 가치를 지니고 있기 때문에 금전적인 이익을 얻기 위하여 이들간의 의도적인 혼입이 유발되기 쉽다. 이들 인삼속 식물들은 현재 재배가 가능한 인삼, 미국삼, 전칠삼, 베트남삼을 제외하면 대부분이 자연상태에서 채취되고 있다. 인삼은 인삼속 식물 중에서 가장 잘 알려진 식물로 상대적으로 가장 다양한 종류의 진세노사이드를 함유하고 있다. 그러나 인삼은 4~6 년의 긴 재배기간이 소요되며 환경 스트레스를 제어하고 고품질의 인삼을 재배하기 위해서는 많은 노력이 필요하다. 따라서 인삼 산업에 전반적으로 영향을 미치고 있는 이러한 문제점 및 한계점을 극복하고 이들간의 유전 다양성과 진화적 유연관계를 이해하기 위하여 7 개 인삼속 식물의 엽록체 유전체를 대상으로 비교 유전체 분석을 수행하였으며 이를 기반으로 하는 종판별 시스템을 구축하였다. 더 나아가 진세노사이드의 대량생산을 위하여 모상근을 이용한 형질전환 시스템을 확립하였다.

첫 번째 단락에서는 5 개 인삼속 식물(인삼, 미국삼, 전칠삼, 죽절삼, 베트남삼)의 전장 엽록체 유전체 서열을 대상으로 비교 유전체 분석을 수행하였다. 5 종 엽록체 서열간의 비교를 통해 길고 짧은 반복서열의 분포를 밝혀내었으며 다수의 Indel 과 SNP 를 발굴하였다. 길이가 긴 Indel 지역들을 대상으로 5 개 인삼속 식물을 구별할 수 있는 분자마커를 개발하였으며 이중 8 개는 종특이 마커로 하나의 종을 다른 종들로부터



특이적으로 구별할 수 있었다. 이들 마커는 안정적으로 작동하며 다형성을 쉽게 파악할 수 있기 때문에 관련 연구 및 인삼 산업 현장에서 유용하게 쓰일 것으로 기대된다.

두 번째 단락에서는 2 개의 인삼속 식물(화엽삼칠삼, 백삼칠삼)의 엽록체를 추가로 더 완성하여 총 7 개의 인삼속 식물의 엽록체 유전체를 대상으로 비교 유전체 분석을 수행하였다. 이를 통해 전체 엽록체 유전체 수준뿐 만 아니라 단백질을 암호화하는 유전자 내에 존재하는 다수의 SNP 를 발굴하였다. 또한 엽록체 유전체를 기반으로 한 유연관계 분석을 통해 7 개 인삼속 식물들과 근연 두릅나무과 식물들간의 진화적 유연관계를 명확히 밝힐 수 있었다. 엽록체와 미토콘드리아 유전체 서열을 서로 비교하여 약 38.6%의 엽록체 유전체 조각들이 미토콘드리아에 삽입되어 있는 것을 밝힐 수 있었다. 따라서 미토콘드리아로 전이되어 있는 유전자 지역과 종내변이를 보이는 지역을 제외한 나머지 유전자 지역으로부터 18 개의 종특이 SNP 마커를 디자인하였다. 이들 마커는 모두 성공적으로 각각의 종을 식별할 수 있었으며, 이는 앞으로 인삼 산업과 관련된 부정 원재료 혼입을 방지하는데 도움을 줄 수 있을 것으로 생각된다.

마지막 세 번째 단락에서는 *Agrobacterium rhizogenes* 을 이용한 형질전환으로 얻은 모상근을 기반으로 효율적인 진세노사이드 대량 생산 체계를 구축하였다. 5 개의 형질전환 된 계통 중에서 가장 형질 전환 효율이 높은 연풍 품종과 강화 재래종을 선발하였으며 *rol* 유전자에 대한 PCR 과 RT-PCR 를 통해 이들 모상근이 *Agrobacterium rhizogenes* 에 의해 유도 되었음을 확인할 수 있었다. 모상근은 부정근보다 더 빨리 자라며 옥신을 처리하지 않아도 성장이 유지된다. 비록 모상근의 진세노사이드 함량은 부정근보다 떨어지지만 몇몇 주요 진세노사이드의 경우 부정근과 비슷하거나 오히려 더 높은 함량을 보여준다. 모상근의 대량 생산에는 바이오리액터 시스템이 가장 적합한 것으로 생각된다.

본 연구를 통해 생산된 유전체 정보는 향후 연구를 위해 유용하게 사용될 수 있을 것으로 기대된다. 먼저 엽록체 유전체 서열을 기반으로 7 개의 인삼속 식물들과 근연 유사종들간의 진화적 유연관계를 밝혔으며, Indel 과 SNP 두 종류의 분자 마커를 이용하여 7 개 인삼속 식물들의 종관별 시스템을 구축하였다. 이는 향후 인삼 산업에서 부정원재료 혼입을 방지하는데 기여할 것이다. 본 연구에서 확립된 형질전환 시스템은 관련 후속 연구를 이어나가는데 도움을 줄 것이며, 형질전환 된 모상근은 진세노사이드 및 인삼의 대량생산을 위해 유용한 재료로 사용될 것이다.

학번: 2014-30833